

Abstract of the Dissertation

Probabilistic and Flux Landscapes of the Phage λ Genetic Switch

by

Nathan Borggren

Doctor of Philosophy

in

Physics

Stony Brook University

2011

The phage λ infection of an *E. coli* cell has become a paradigm for understanding the molecular processes involved in gene expression and signaling within a cell. This system provides an example of a genetic switch, as cells with identical DNA choose either of two cell cycles: a lysogenic cycle, in which the phage genome is incorporated into the host and copied by the host; or a lytic cycle, resulting in the death of the cell and a burst of viruses. The robustness of this switch is remarkable; although the first stages of the lysogenic and lytic cycles are identical, a lysogen virtually never spontaneously flips, and external stressors or instantaneous cell conditions are required to induce flipping. In particular, the cell fate decision can depend on the populations of two proteins, *cI* and *Cro*, as well as their oligomerization and subsequent binding affinities to three DNA sites. These processes in turn govern the rates at which RNAP transcribes the *cI* and *Cro* genes to produce more of their respective proteins.

Although the biology in this case is well understood, the fundamental chemistry and physics underlying the bistability remains elusive. In this work, a dynamical model of the non-equilibrium statistical mechanics

is revisited, generalized, and explored. The low number of proteins and other sources of noise are non-negligible and corrections to the kinetics are essential to understanding the stability. To this end, general integral forms for advection-diffusion equations appropriate for finite element methods have been developed and numerically solved for a variety of mutants and assumptions about the state of the cells. These solutions quantify the probabilistic and flux landscapes of the ensembles' evolution in concentration space and are used to predict the populations of the cell states, entropy production, passage times, and potential barriers of wild type and mutant bacteria to illuminate some structure of the configuration space from which Nature naturally selects.

Contents

List of Figures	v
List of Tables	vii
Acknowledgements	x
1 Introduction	2
1.1 Computation, Noise and Limit Cycles in Biophysics	2
1.1.1 Chemical Oscillations	10
1.1.2 Dynamics	10
1.1.3 Diffusion	11
1.2 Computers and Phage λ	11
1.3 phage λ	12
1.4 Evolution	17
1.4.1 Natural Selection from...	17
1.4.2 Natural Selection for...	18
1.5 Three modes of Computation	18
1.5.1 Binding Sites	18
1.5.2 Polymerization	19
1.5.3 Solitons	19
1.6 Methods of Numerical Analysis	20
1.6.1 FEniCS	22
1.6.2 ROOT	22
1.6.3 Aleph	22
1.7 Overview	22
2 Binding Polynomials and the Chemistry of the O_R Region	24
2.1 ΔG Measurements	24
2.2 A Partition Function; RNAP or no RNAP?	28
3 Dynamics Near Equilibrium	31
3.1 Gene Expression and General Dynamical Systems	31
3.2 Stationary Points	35

4	<i>Diffusion; Evolution of Ensembles</i>	39
4.1	Master Equations, Langevin Dynamics, or Diffusion Equation	40
4.1.1	An Integral Approach for Diffusion Equations and Passage Times	40
4.1.1.1	Diffusion Equations	40
4.1.2	On Convergence	42
4.1.3	Choice of Gauge	42
4.1.4	$\Delta x \Delta J$	43
4.2	Passage Times	44
4.3	Noise-induced stability	44
4.4	Little's Mutants	47
4.4.1	Wild-Type Evolution	47
4.4.2	Steady state properties for O_R Mutations	47
4.5	Hybrid particle and ensemble approaches	48
4.6	Variations on a theme	48
5	<i>Conclusions and Future Directions</i>	51
5.0.1	On Computation and Information	52
5.0.1.1	The Physical Interface	52
5.0.1.2	The Human Interface	53
5.1	On the importance of generality and the abstract	54
5.2	On the importance of the particular and the measurable; Case Studies	55
5.2.1	Robustness, robustness, stability, and stability	57
5.3	Algebraic, Analytical, or Computational Methods?	57
5.3.1	Non-Associative Commutative Algebra	58
5.3.2	Recursion Equations for Expansion coefficients	58
5.4	From biology to physics to biology	60
	References	69
A	Appendix; Tabulation of Results	70
A.1	Results	70

List of Figures

1.1	λ phage requires a host for RNAP and ribosome to express its genes.	7
1.2	The two <i>E. coli</i> outcomes, lysogenic above and lytic below, of a λ phage infection. The O_R is yellow, the active cI gene is blue, the active Cro gene is red, and grey are other active genes in that stage of the cell cycle. The Att gene in green is the gene responsible for incorporating the virus into the bacteria. White regions are DNA segments inactive during that stage.	8
1.3	Some basic features of the λ phage genome.	13
1.4	Zooming into the O_R site reveals more structure, the P_R site overlaps with O_{R1} and O_{R2} and P_{RM} overlaps with O_{R3}	14
1.5	Depicted are three possible states of the O_{Ri} site. Each of the three DNA sites can be either empty, occupied by a cI dimer, or occupied by a Cro dimer.	15
4.1	Finite elements construct a surface by combining simplices together. The quadratic elements used fit the function inbetween the mesh points with a quadratic formula.	42
4.2	This would be an example of a legitimate python code that runs indefinitely. Computationally, the asymptotics of the equations require a function that does not terminate.	43
4.3	Calculating expectation values is straightforward, even on a command line.	43
4.4	The first passage time with an absorbing boundary condition within 10 units of the origin. The origin is extremely unstable; The steepness of the funnel of the first passage time landscape illustrates the overwhelming propensity for the system to be driven towards some non-trivial state. The numbers, $\approx 10^8$, appear in seconds.	45
5.1	Variational forms used in this work.	53

5.2	This is the \aleph_{00000} Gui demonstrating user interaction with equations of motion. It reads equations of motion of a system and builds a Gui to interact with that system	53
A.1	Contours with stationary Ci and Cro production with O_{R_1} fixed. The axis are in protein numbers. 3 of the 9 mutations are destabilized: $\lambda_{112}, \lambda_{122}$, and λ_{132}	71
A.2	Contours with stationary Ci and Cro production with O_{R_1} fixed. The axis are in protein numbers. 3 of the 9 mutations are destabilized: $\lambda_{212}, \lambda_{221}$, and λ_{231}.	72
A.3	Contours with stationary Ci and Cro production with O_{R_1} mutated to O_{R_3}. 5 of the 9 mutations are destabilized: $\lambda_{311}, \lambda_{321}$, λ_{322}, λ_{331} and λ_{332}.	73

List of Tables

2.1	Values of variables referenced by the partition function.	27
4.1	Values for diffusion tensors. λ_{323p} in Little et al. [1999] is nearly λ_{222} , the <i>cI</i> affinities are identical and <i>Cro</i> values differ by -.6 kcal/mol in the first and third site.	47
4.2	Values of variables referenced by the partition function with RNA polymerase states included. A.22 uses the values of $[RNAp]$, ΔG_{30} , and ΔG_{03} listed first, whereas A.23 and A.24 use the values listed second.	50
A.1	Mutants with one stationary point.	74
A.2	Mutants with two stationary points	74
A.3	Mutants with three stationary points	75
A.4	Mutants with three stationary points (continued)	76
A.5	The time evolution of a gaussian initial condition relaxing into the steady state for the wild type. The higher peak corresponds to lytic while the wider peak is lysogenic.	77
A.6	The time evolution of a gaussian initial condition relaxing into the steady state for the 121 mutant. The dominant peak is the lysogenic peak.	78
A.7	The time evolution of a gaussian initial condition relaxing into the steady state for the 323 mutant. The dominant peak is lytic. The gap is reduce compared to wild type.	79
A.8	The time evolution of a gaussian initial condition relaxing into the steady state for the 323p mutant. The dominant peak is lyso-genic.	80
A.9	Properties of mutant λ_{121}	81
A.10	Properties of wild-type λ_{123}	82
A.11	Properties of mutant λ_{131}	83
A.12	Properties of mutant λ_{133}	84
A.13	Properties of mutant λ_{211}	85
A.14	Properties of mutant λ_{213}	86
A.15	Properties of mutant λ_{222}	87

A.16 Properties of mutant λ_{223}	88
A.17 Properties of mutant λ_{232}	89
A.18 Properties of mutant λ_{233}	90
A.19 Properties of mutant λ_{312}	91
A.20 Properties of mutant λ_{313}	92
A.21 Properties of mutant λ_{323}	93
A.22 The time evolution of a gaussian initial condition relaxing into the steady state for the wild type with RNA polymerase in the partition function and constant diffusion tensor. Notice that the two states, lytic and lysogenic, are gapped only for the window of time from 4800 to 8400 seconds. Although the emphasis in the literature is with steady states, it is an interesting possibility that the two state phenomena is a strictly out of equilibrium phenomena, existing only for a window of time in the cell. A bacteria cell cycle is around 1800 seconds, suggesting that for many distributions the steady state is not reachable in the necessary time scale.	94
A.23 The time evolution of a gaussian initial condition relaxing into the steady state for the wild type with RNA polymerase in the partition function and constant diffusion tensor with different parameters from A.22 . The lysogenic peak in the classical dynamics is near 1500 cI proteins. To find a distribution with a second peak required this enlarged domain and was facilitated by an initial condition with many cI (3500). The short axis is 0-1000 Cro proteins, and the long axis is 0-5000 cI proteins. One can see the global effects through the large displacement of the lysogenic peak from its stationary point.	95
A.24 The time evolution of a gaussian initial condition relaxing into the steady state for the wild type with concentration-space dependant diffusion. The peak corresponds to the lytic state. Two strong tails develop, one extending towards potentially another peak.	96
A.25 Entropy and Energy densities evaluated along example paths for mutations in $O_{R2,R3}$ with the O_{R1} fixed. Different trajectories can have different limiting values of entropy and energy density depending on the path taken.	97
A.26 Entropy and Energy densities evaluated along example paths for mutations in $O_{R2,R3}$ with the O_{R1} mutated into O_{R2} . Different trajectories can have different limiting values of entropy and energy density depending on the path taken.	98

A.27 Entropy and Energy densities evaluated along example paths for mutations in $O_{R2,R3}$ with the O_{R1} mutated into O_{R3}. Different trajectories can have different limiting values of entropy and energy density depending on the path taken.	99
--	-----------

Acknowledgements

The original title of this thesis, *One Continuous Mistake*, seemed to have already been taken, though time has now evolved sufficiently long to have absorbed enough contribution from elsewhere to break this continuity sufficiently to warrant a new name and all of the residual mistake is entirely my own.

I would like to thank Jin Wang for this opportunity, guidance in looking universe, and for the space and time to diffuse through many ideas and approaches to the questions he had asked of me. His grand canonical ensemble: Wei Wu, Weixin XuBo Han, Zhiqiang Yan, Yuan Yao , Xiaosheng Luo, Haedong Fong, Jeremy Adler, Qiang Lu, Zaizhi Lai, Ruonan Lin, Ronaldo Oliveira, and to those whose bodies have fluctuated elsewhere but kept their influence close: Liu Fang, David Lepzelter, Saul Lapidus, Keun Young Kim, and Amber Carr have been endlessly knowledgeable, kind, and helpful and add some reason to the rhymes and rhyme to the reasons to a vast array of problems in the field.

I would like to thank my committee for everything, and in particular it is quite a privilege that I have been able to ponder deeply questions with Ajit Nuggenhali about nuclear matter and Dmitri Averin with condensed matter. This thanks extends to Jiangyong Jia, Roy Lacey, Rui Wei, Arkadij Taranenko, John Anderson, Damian Reynolds and Alex Mwai as well as the rest of the Phenix collaboration, Vasili Semenov, Jie Ren, and Supradeep Narayana have been generous with their time, expertise, laboratories as well. My other FEniCS collaboration has been fast, thorough, and efficient with their help and expertise on the rare occasions direct mimicry of their methods didn't suffice. The scientific influence of Chris Tartamella, Stanislav Srednyak, and John Schreck are also deeply ingrained in this work, but that influence is trivial next to their friendship.

I would like to thank Mark Ptashne for an early version of [Ptashne \[2011\]](#) and Xue Lei and Ping Ao for helpful discussion of units.

I have never stopped counting in densities since leaving Roger Yelle's company though his teachings and insight still seem to anticipate me and welcome me in every new direction I turn.

My mom, dad and E^3

and Amber, la luz, *Let simple and old fashioned myself stay with you, while ordinary things go disappearing from this world.*

1 *Introduction*

Life is an entanglement of lies to hide its basic mechanisms.
William S. Burroughs, *The Place of Dead Roads*

1.1 *Computation, Noise and Limit Cycles in Biophysics*

It is a season in which the porousness of the borders delineating the subjects that comprise Knowledge are growing more so, or altogether vacuous. The overlap is revealing some truths to be quite universal. It has been said that *more is different* [Anderson \[1972\]](#) and *information is physical* [Landauer \[1961\]](#), and concepts involving emergence and entropy are proving to be fundamental to understanding the physical basis of biological systems. The notion of a complete and clockwork deterministic universe has been left behind, and the necessity of dealing with probabilities, partial information, measurement, and open systems has become essential to all of the major sciences.

Among these disciplines, the common goals have become those of elucidating which states are possible, determining the relative probabilities or amplitudes of those states, and examining their stability and robustness. In the subject of physics, the need for these particular goals is most easily evidenced by the existence and necessity of quantum mechanics, which assigns to each system a wavefunction from a spectrum of permitted states; entanglement amongst these states; and, by measurement, results ultimately in the collapse of the wavefunction to a single eigenstate of the spectrum. In chemistry, it has become increasingly necessary to delineate the single pathway that a molecule may take, as it selects a seemingly random walk in a solvent, or geometrically aligns with a catalyst in a reaction. In biology, evolution provides a framework by which we understand how each member in a myriad of diverse species becomes finely adapted to their single, specific niche, and how these populations are interrelated.

Although the underlying theories are generally accepted, there remain open questions and challenges in the implementation of these theories that are similar across disciplines. Analogies can even be present in the mathematical formalisms that are often developed in parallel for these subjects. For example consider time evolution in quantum mechanics, Brownian motion, and population dynamics. The integral of a specified initial condition is commonly implemented iteratively as a matrix operator acting on a vector which represents the quantum state in the case of quantum mechanics, the population densities in the case of chemical

dynamics, or the population density in the case of an ecosystem.

The limit cycles are of considerable interest. Although teleological explanations in evolution are largely vanquished [Mayr \[2002\]](#), at least some of the properties of living matter can be predestined by the highly non-negligible physical and chemical constraints placed on it by the supporting planet. Understanding physically how systems acquire their properties through gene expression and feedback from an environment leads to fundamental physics. However large the number of unknown degrees of freedom, they serve as a source of energy and entropy for the system, or as a place to deposit energy and entropy. In some shape or form, this exchange introduces a mechanism for choice, chance, or randomness into the time evolution of the system, which is commonly implemented as a matrix acting on a vector.¹ Be it the decoherence processes that select an eigenfunction out of entangled quantum states or the natural selection processes that select the fittest of a species, it is *time* that selects an actuality out of the sea of possibilities that could potentially occur. That is to say, whatever the equations, it is the asymptotics of those equations that are of interest. When the system is open, the nature of this selection speaks equally as much of the measurement device used or the planet underfoot as it can of the system measured or the species selected.

Assuming that we have equations that describe these phenomena, is it feasible to imagine finding solutions to them? The question we ask is what the best models are for the physics we can develop given the computational resources that are available, and how we can improve the precision of the physics as our algorithms and computing power improve. When we point our computers at the abstractions we use for these systems, the polynomial algorithms seem nowhere to be found. While for biological problems, even by counting only the emergent degrees of freedom (proteins) of the emergent degrees of freedom (RNA) of the emergent degrees of freedom (DNA) of the emergent degrees of freedom (the nucleotides) of the emergent degrees of freedom (the periodic table) of the emergent degrees of freedom (the protons and the neutrons), our ability to directly probe the relevant scales of length and time is limited by our current information technology.

Because many questions in one field are categorically identical to questions in another, the achievements and methods in one science are non-negligible to the others. Often they are the same achievement or require cooperation. The contribution of physics thus far to biology is largely that of engineering: it is the microscopes, spectrometers and transistors that are mainly of service. There is no evidence suggesting that the physics of biology requires more than quantum mechanics, statistical mechanics, and classical mechanics to operate, but the extent to which the solutions to our equations lag behind our equations provides ample room for doubt and countless theorems remain to be demonstrated to *show* this.

¹Consider, for example, $\dot{x} = Ax$, $i\hbar\partial_t\Psi = H\Psi$, $\partial_t P = -\partial_i(D_{ij}\partial_j P - F_i P)$ with solutions, $x(t) = e^{At}x(0)$, $\Psi(t) = e^{-\frac{i}{\hbar}\int_0^t H dt}\Psi(0)$, and $P(t) = e^{\int_0^t (D_{ij}\partial_j - F_i) dt}P(0)$ respectively. The essential problem, be it a continuous Markov process described by A, Schrödinger's equation, or the Fokker-Planck equation, the equation boils down to the evaluation of an exponential of an operator.

As an example, biology has provided us with the ingenuity through experiments to deduce the relevant underlying networks for the λ phage Ptashne [1992]², the flexibility of those networks Atsumi and Little [2006], Little et al. [1999], and how to introduce and parameterize the empirical force fields that are introduced to study the dynamics Shea and Ackers [1985]. These numbers are improved every day as better microarray data is accumulated and understood³.

Evolution, like gravity and thermodynamics, is so seamlessly entangled into our lives as to often go unnoticed. Once, however, the concept is understood it is quite self-evident, and is as readily seen as the selection of an item off of a menu or the selection of students or employees from a pool of applicants. This interplay between chance and determinism, the arbitrary and the necessary, and populations and their individuals provides more than enough room for complexity to emerge, even before granting a capacity for volition to the individuals.

The key mechanism again is emergence - be it quarks to form a proton, nucleotides to form DNA, or cells to form an organism, the salient necessity for evolution to occur at any hierarchy of a universe is a mechanism to build up systems of arbitrary complexity out of simpler and readily available resources from the nearby hierarchical levels Jacob [1977], McCoy [2010]. Unsurprisingly, we find ourselves present in such a universe, and Nature has no shortage of mechanisms to build complex systems, and clever algorithms for their improvement Darwin [2011], tol [2011]⁴.

It has been said that the universe *is* a quantum computation Lloyd [2006], and it will not resist being perceived as one, but other modes of computation the universe readily performs, those using the counting of molecules, are worthy of deep reflection Bennett [1982]. These classical modes can also supply the speedup to which quantum computers esteem. This Brownian computer, as Bennett calls it, is not dissimilar to a quantum computer, in the sense that it relies on interference to decrease the probability of certain possibilities and increase the probabilities of others. The fundamental difference is in the nature of the interference, which for a quantum computation occurs at the amplitude level, where numbers can be complex, and for a Brownian computer this occurs at the probability level, where the numbers are strictly positive. The fundamental similarity between the two modes of computation is that they occur essentially in parallel.

To illustrate, consider an impenetrable door, nay, there is an alphabet with four letters and one password in 4^n words that will let us in were we to enter that word. The keepers of the door also granted us a classical, quantum, and chemical

²and references therein

³In a just universe, the models such as discussed in this thesis would be coupled with data. The rates used in the partition functions should be fit simultaneously with the numbers used in the diffusion tensor as well. We do not take such a course here but provide the reference Kinney et al. [2007], in addition to Shea and Ackers [1985] to illustrate how ΔG numbers are accumulated

⁴The notion of species-as-algorithm is evident in Toffoli [1982], where he gives the example of a typical control function: the octopus.

keypad. The classical keypad takes too long, qubits have their own agenda at those numbers, so how then will we ever open the door?

What is a chemical keypad? Say we have n sites and we enter our code by binding 1 of our 4 molecules onto each site, and if we get the arrangement just right, then the door opens. A molecule will simply go away after some time if it did not fit.

Our only hope then without any more prior information is Bennett's Brownian computer. We simply gas the keypad with concentrations of all the molecules and wait. Sooner or later the necessary configuration will occur and our key is found and the door is opened.

This is still a classical computation in the sense that it is no different than our fingers punching numbers into a keypad as far as the information processing is concerned. Indeed the speedup here is simply of scale, as molecules can typically bind and unbind to sites and each other faster than we can push buttons. The probabilistic nature of this process comes only as a result of our description of the scenario. We have only macroscopic information, the thermodynamic variables, and may be able to tune them to some degree, but the goal is a particular conformation. This information is not accessible to thermodynamics proper, but we know the answer once the goal has been achieved, as we read it out of the keypad only after the door is opened. As is often the case, the key itself might fall out of the door and this information too is lost, but sometimes just an open door is sufficient to proceed.

Evolution *is* such a mode of computation: a wide variety of possibilities are attempted, and if any open doors, or are built to last, then doors open and things last. The rate and mechanisms it achieves to do this provides curious insight into new computers and algorithms for achieving and dealing with computational complexity, typically involving non-binary modes of thinking.

Granting a billion years from the birth of Earth to the birth of the first prokaryotes, say 4×10^6 bp (base pairs) Mayr [2002], it can be inferred that the effective rate at which natural selection processed the $4^{4 \times 10^6}$ configurations is $4^{4 \times 10^6 - 27}$ Hz. This is a rephrasing of Levinthal's paradox in protein folding for the problem of evolution. It is thus safe to infer that natural selection is not picking numbers at random from a flat distribution.

Genes are an example of an algorithm that has provided exponential speedup to this sampling. Genes might be a transcript for a protein and take up a thousand nucleotides of space, and at this stage Nature's trial and error processes of evolution revolve around a sampling of the space of gene combinations; this is a considerably truncated space from the space of permissible DNA chains. Sexual reproduction is one example exhibiting the benefits of such a sampling, and it emerged as a means of speeding up the quest for fitness about 2.7 billion years ago. The DNA of the resulting offspring is far removed in a symbol space from either parent's DNA, yet the genes persist. It is quite a luxury that a child should take after their parents.

With a configuration space so vast and the demands of a niche so extreme, one

can hardly suspect a few random bits flipping to send one to a more favorable potential well. Quite the contrary is fact and indeed the structure of DNA, with its Watson-Crick pairing and strength of its bonds, is largely preventative of just this situation. Unsurprisingly, there exist proteins and mechanisms that specialize in DNA repair to further parry these crude attempts to debase the bases. In eukaryotes, DNA is further protected via a profound capacity for packing and elaborate protection from histones and proteins.

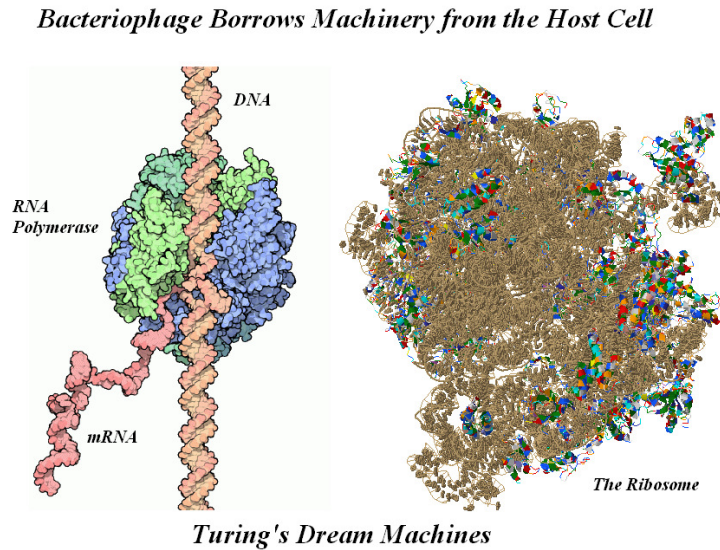
So it is genes that drive development and the space of gene configurations that DNA prefers to play trial and error with, for exactly the reason to exponentially speedup the search for fitness. There is then a very strong and productive motive to treat the genes and their functions as the relevant degrees of freedom, and indeed that is the approach and success of systems biology. [Ptashne and Gann \[2002\]](#), [Nüsslein-Volhard \[2006\]](#)

It can be seen that evolution can proceed with large steps. There are mechanisms in which the step nature of evolution of DNA itself is evident. This revolves around the discussion of genes, not nucleotides, where very specific and relatively long transcripts can be inserted irreversibly into a genome. Such is the case we study now, the phage λ infection of an *E. coli*. The biological content of this thesis is well known and actively extended, [St. Pierre \[2009\]](#), [Ptashne \[1992\]](#).

λ is not alive and requires an *E. coli* host cell to provide the habitat and ingredients, in particular RNA polymerase and the ribosome, to express and replicate the genes present in that genome. A head provides protection to the DNA outside of a cell as the virus uses a tail to seek out a host upon which it pierces the cell membrane and injects the DNA into the cell. A special feature of the genome is the six bases on each end; they are unpaired 'sticky ends,' and are dual to each other. These sticky ends stick to to each other upon entering a cell and form a ring. The viral DNA then waits for the cell contents to regain their closest approach to equilibrium. This tricks the RNA polymerase and ribosome in the host into expression and synthesis [1.1](#) of its genes.

This splitting of a cell population into different possible fates can be seen in a variety of circumstances. Some cells burst into a multiplicity of phages immediately (lysis), while the bacteria that do survive are irreversibly transformed into a new genome, now nearly 48,502 bp longer, and henceforth replicating along with its old identity and the entire phage (lysogeny), see [1.2](#). This behavior seems more akin to a jewel heist than a thermal, stochastic, or a quantum fluctuation. It is believed that noise influences the decision at the level of selection between fates. This noise has been studied elsewhere, [Zhu et al. \[2004\]](#), [Wang et al. \[2010\]](#), [Arkin and Ross \[1998\]](#), but the exact nature of this noise is not understood, and is typically presumed Gaussian for computational and analytic convenience. The methods developed to study this noise can easily implement more complicated noise, for example, as known functions of space and time. This switch is a perfect example of a switch component of a Brownian computer, but in the end just one component of a larger λ gene network, presented in [Arkin and Ross \[1998\]](#), [St. Pierre \[2009\]](#).

Figure 1.1: λ phage requires a host for RNAP and ribosome to express its genes.



The rates at which these transcriptions occur enter the model's equation of motion as simply numbers and set the time scale of the computation, i.e. the expression of the gene, so let us revisit RNAP, and the ribosome, to first marvel at two of the most amazing molecules this side of the big bang, and chieftains among the hyperspecialized computational hardware inside a cell.

Indeed it was RNA polymerase that was the source of inspiration for the Brownian computer in the first place, and its functioning had been described in information theoretic terms already in 1982 [Bennett \[1982\]](#):

To see how a molecular Brownian computer might work, we first consider a simpler apparatus: a Brownian tape-copying machine. Such an apparatus already exists in nature, in the form of RNA polymerase, the enzyme that synthesizes a complementary RNA copy of one or more genes of a DNA molecule.

Bennett goes on to describe the RNA polymerase action in terms of a one-dimensional random walk along a DNA chain. In the forward reaction RNA polymerase opens the DNA to read the content and append to the growing RNA strand the nucleotide part of the relevant nucleotide pyrophosphate ATP, GTP, CTP, or UTP and subsequently releasing the pyrophosphate molecule into the cell, advancing a notch on the DNA chain. In the backward reaction RNA polymerase takes a free pyrophosphate (PP) from the cell and combines it with the last RNA bit and releases into the cell the ATP, GTP, CTP, or UTP. In equilibrium this reaction would balance and RNA and protein synthesis would not occur. However, the metabolic processes are such that a bias in the direction of RNA synthesis occurs by removing free PP and supplying ATP, GTP, UTP, and CTP to the vicinity of RNA

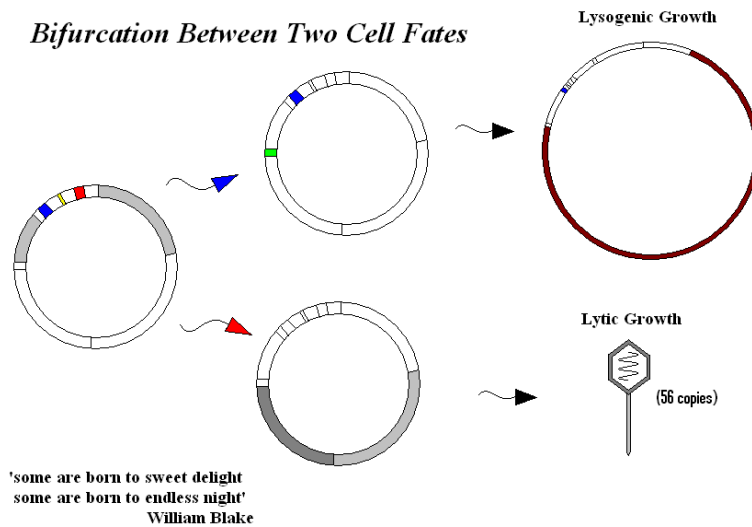


Figure 1.2: The two *E. coli* outcomes, lysogenic above and lytic below, of a λ phage infection. The O_R is yellow, the active cI gene is blue, the active Cro gene is red, and grey are other active genes in that stage of the cell cycle. The Att gene in green is the gene responsible for incorporating the virus into the bacteria. White regions are DNA segments inactive during that stage.

polymerase. The numbers used in his analysis are RNA polymerase advancing at 30 nucleotides/s (330 forward reactions -300 backward reactions), and dissipates 20kT per nucleotide all while making less than one mistake per ten thousand nucleotides. [Bennett \[1982\]](#)

By 1999, this one-dimensional random walk was directly observed in single molecule experiments [Guthold et al. \[1999\]](#). The rates therein were much lower (1.5 nucleotides/s \pm 0.8 nucleotides/s), but these processes were likely slowed down by the experimental conditions which confined the DNA to diffuse along a 2D substrate, restricting the motions of many degrees of freedom, to facilitate the observation of RNA polymerase moving along the DNA as if it were a 1D chain. It had been argued that effectively diffusing in lower dimensions enables RNA polymerase to find the promoters much more easily.

λ has developed an additional strategy that speeds up this transcription involving cI protein [Ptashne \[2011\]](#). It involves a relatively small cooperativity (-2kcal/mol) binding between a cI dimer and RNAP at the promoter site, which also shortens the time RNAP requires to diffuse to that point. This is a small number and can not alone describe the observed factor of 10-100 increase in transcription rate when cI is present. An explanation I provide here is that this speedup could be a result of another complex binding process, where cI octamers short the DNA loop, as has been observed. That this results in a speedup of transcription is a result of the simple fact that, if RNAP can maintain a 30 nucleotide/s rate, this

rate is increased in units of m/s if the chain is shortened and there is simply less distance to diffuse to the end of the transcript and back to the promoter. Indeed the looping between the O_R and the O_L leaves RNAP again approaching in physical space the promoter as it is finishing the transcription of cI . This increase does not occur for the Cro gene since the shorting of the loop does not contain the Cro gene, only the cI gene, and requires cI to be occupying the promoter site for Cro anyway.

The subsequent RNA strand is fed to the ribosome which implements a many-to-one dictionary, converting three consecutive nucleotides into an amino acid. The 64 4-letter words of length 3 get mapped to 20 amino acids, aiding in the formation of the protein. This allows for a large multiplicity of DNA words that can create the same protein and thus allow for a finer resolved selection process to occur amid the words that produce a given protein.

cI is a chain of 236 amino acids and cro is 66. The rates at which these processes occur are actively measured and correspond to the value of ν_{xi} for cI and Cro in this situation. The rates for the lambda phage production of cI and Cro used in these simulations are 0.115 cI monomers/second for states with cI assisting RNAP, and 0.30 Cro monomers/second. Other impacts on the rates are the 0.01045 cI monomers/second contribution from certain configurations, and that the ribosome produces 20 Cro for each reading of the Cro transcript. These numbers are chosen to correspond to the values use in [Zhu et al. \[2004\]](#), [Wang et al. \[2010\]](#) where this dynamical system is investigated. The complete set of parameters used in this simulation will be shown in the dynamics section.

There are two sources of non-linearity in the equations of motion, first from there simply being multiple binding sites, but the main complication, and source of error in this work, is that both cI and Cro first dimerize in the cell and bind in dimer form to DNA with a helix-bend-helix motif. Different configurations have slightly different affinities and thus occur with different frequencies. All the configurations however have negative ΔG and as such are spontaneous reactions; they are energy favorable in comparison with an empty strand of DNA. In addition, cI dimers are found to bind cooperatively on adjacent sites. This added energy bonus from cooperation aids in the competition with Cro. Cro is also limited by the fact that both O_{R1} and O_{R2} site need to be unoccupied for RNAP to bind to the Cro promotion site. The partition function is dominated by the triply liganded species and the faster rates and large bursts for Cro get largely repressed by the relative scarcity of the singly liganded states in comparison to the doubly liganded states that promote cI .

The phage genome in a lysogen patiently waits for cell conditions to appear, a process known as induction, that switch the genes being expressed to phage-bearing ones and result in bacteria death and phage progeny. Ultraviolet light, small cell volume, and elevated Cro or depleted cI protein levels are known to be able to induce the switch. It is the object of this dissertation to further quantify the nature of this swith and compare with the words nearby in simple space. These are very specific mutations in the O_R , chosen so that the binding properties can

be predicted, and switch properties studied. A result of this thesis, and a testable hypothesis, is that energetically the use of ultraviolet light for induction is a little extreme, as radiation targeted at the O_R equivalent to $15kT$ should be sufficient to provide the energy for some of the lysogens to flip. Indeed, NMR techniques specialized to control cI to the extent that it can be removed from O_{R1} and O_{R2} should also be sufficient for induction even at larger concentrations.

1.1.1 Chemical Oscillations

That classical chemical reactions used to describe the balancing of morphogens within an embryo is sufficient to give rise to non-trivial spatial and temporal structures was noted carefully in the last works of Alan Turing [Turing \[1952\]](#). See [Kondo and Miura \[2010\]](#) for a modern review.

Classical treatment of chemical reactions is sufficient to understand many rhythms of the body, like a heart beating, and it is wise to push the utility of such treatments as far as possible.

Reactions are not even necessary for oscillations to occur, indeed mixtures of perfect gases oscillate on grand scales in atmospheres in an attempt to maintain hydrostatic equilibrium for the mean density despite tidal or other forces [Niemann et al. \[2005\]](#), [MÃijller-Wodarg et al. \[2006\]](#).

The variety of chemical oscillations considered in this thesis are particular to those believed to be relevant in gene expression, which involve the binding of ligands to operator sites resulting in the creation of more ligands.

The chemistry section of this work states quantitatively our basic assumptions: namely the partition functions used to describe the equilibrium statistical mechanics of the protein/DNA reactions. It is stated to a reasonable generality what shapes these partition functions can make, facts that we will need to motivate the extent of accessible dynamics to which these binding polynomials along with transcription factors can give rise.

At low numbers, chemical oscillations take on much greater complexity, where noise can play dominating roles. Most importantly this work relies on the discussions in [Wang et al. \[2008\]](#).

1.1.2 Dynamics

Universal computation takes on a whole new meaning in chemical and quantum computation. One need not seek only classical answers to classical questions, as analog circuitry and quantum computation can greatly enrich the classes of problems that can be understood. The dynamics resulting from gene expression could be of interest to the purest of mathematicians, and in some sense are the n -body generalization of Hilbert's 16th problem, which remains unsolved. We applaud chemistry and biology for their courage in asking such questions. Hence it is quite appropriate to begin with the most general equations for dynamical systems. The qualitative and quantitative theories of dynamical systems are old

and advanced [Liapounoff \[1893\]](#), [Volterra \[1931\]](#), [Arnold \[1978\]](#), [Perko \[2001\]](#)⁵ and the need for general methods to general equations in simulating gene expressions is noted in [Zhu et al. \[2004\]](#), [Wang et al. \[2010\]](#), [Aurell and Sneppen \[2002\]](#).

The codes used for this work were written to maintain this generality and the software can be used to study dynamical systems over real or complex fields.

1.1.3 Diffusion

Of the few equations Schrödinger reluctantly allows appear in *What is Life?* [Schrödinger \[1992\]](#) the first is

$$\partial_t P = D \Delta P \quad (1.1)$$

Which he need not trouble you to explain. The gradients and diffusion of which he speaks are in physical space, whereas we will be concentrating on concentration space. The motive and solutions are, however, similar.

The robustness of the phage λ is of the type Schrödinger would have claimed at the time to be impossible. He deliberates extensively on the $\sqrt{(n)}$ law and expected that where there is order, such as ferromagnetic order, there will be a requirement of a large number of particles involved. Yet the order one would expect from large concentrations is still present at the small numbers of 100-300 proteins and has given rise to the so-called stability puzzle [Aurell et al. \[2002\]](#).

The analogous equation of (1.1) for the case of concentration space, to which numerical solutions herein are devoted to address, is

$$\partial_t P = \partial_i (D_{ij} \partial_j P - F_i P). \quad (1.2)$$

The codes allow for F_i to take on the generality prescribed in the dynamics section and D_{ij} can be functions of space and time. Our mathematical analysis will proceed with the simplest case of a positive and diagonal diffusion tensor.

A main contribution of this thesis is the adaptation of (1.2) for use with finite elements.

1.2 Computers and Phage λ

Before we posit equations and automate solutions let us note that biological systems are not dissimilar from computers and their behavior is not dissimilar from a program. There is no hope of putting it more eloquently than [Rothstein \[1982\]](#):

“Computers and living systems share many characteristics: they behave selectively, show tremendous differences in response to similar

⁵a study of dynamical systems, to which noise is largely a perturbation, can on no accounts be trivialized or neglected. Many texts on the fundamentals are available, in particular [Perko \[2001\]](#) illustrates the linearization methods to understand the local, qualitative dynamics of a system.

inputs, and are nonequilibrium systems, generally metastable. They usually require an ongoing dissipation to maintain their characteristic behaviors, which depend on stored information. Both involve loosely coupled subsystems, use elaborate internal communication and control systems, and key interactions within them and with their environments are typically irreversible and *all or nothing* (nonlinear).'

One may also add that they have already been selected from a vast space.

The phage λ is simply one of the 4^{48502} molecules DNA could have chosen from this vast space. Yet, this sequence warrants its own name, for it is uniquely adapted to secure the fidelity of that same sequence through the infection of a host cell. To do so, it chooses selectively between lysis and lysogeny and requires an ongoing interaction with RNA polymerase and the ribosome to achieve its steady state protein concentrations through the transcription of the *cI* and *Cro* genes. This involves the coupled subsystems, the *cI* and *Cro* proteins and the O_R , with elaborate internal communication through oligomerization processes, and key interactions within them are indeed nonlinear. The fate of the decision is thermodynamically irreversible. Let us elaborate.

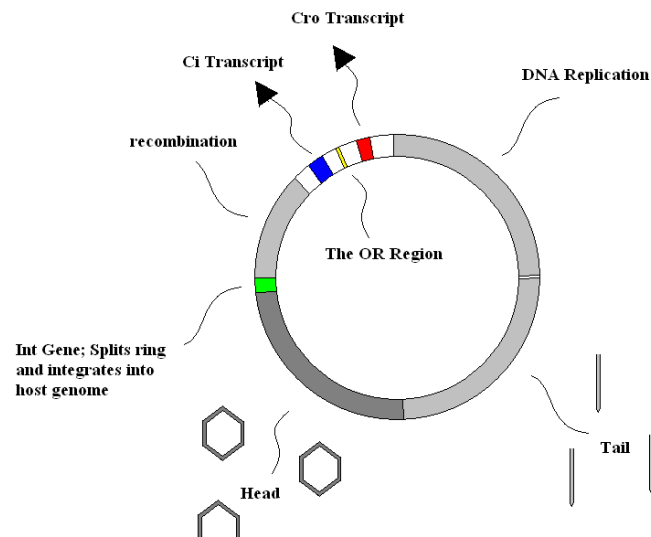
1.3 phage λ

The enterobacteria phage λ infection of *Escherichia Coli* is proving to be an insightful example in biology of how the small-scale molecular interactions involved in gene signaling and expression lead to distinct outcomes [Ptashne \[2011\]](#), [Ptashne and Gann \[2002\]](#). In this case, these outcomes are the very life and death of the cell. A history of ingenious experiments has revealed the interactions involved in this determination of cell fate in exacting detail [Ptashne \[1992\]](#), and the λ phage's strategy for reproduction is a clear example where the molecular strategy involved in the evolution and natural selection of the organism are well understood. The questions then remaining are the chemical and physical questions in the cellular environment, particularly, what the interactions between these macromolecules are, and the type of model that is best used to quantify these interactions. The timescales involved in the lambda phage genetic switch are on the order of 10^2 s to 10^4 s, thereby placing the system beyond the limits of an all-atom molecular simulation with an empirical forcefield. Phenomenology has been introduced, and the origin of the model herein is the formulation of [Shea and Ackers \[1985\]](#) with subsequent revision in [Aurell et al. \[2002\]](#) and propagated by [Wang et al. \[2010\]](#), [Ao \[2004\]](#) and in the framework of [Wang et al. \[2008\]](#): the diffusion equation. This has been adapted herein for the finite element method.

The first description of the λ phage system as a two-state switch was given by the biologists Lwoff, Jacob, and Monod [Lwoff \[1972\]](#). The λ phage is a virus that recognizes and binds to its host, the bacterium *Escherichia Coli*, and subsequently injects its genetic material into the bacterium in the form of double-stranded DNA,

forming a ring. After the virus has infected the cell, there are two main mechanisms by which the viral DNA may be replicated: the lytic cycle, or the lysogenic cycle. In the lytic cycle, the bacterial DNA is destroyed, and the virus takes control of the metabolic activity of the cell with the goal of replicating its own DNA. The host cell is quickly overcrowded with new viruses, and the cell bursts, releasing the new progeny viruses. In the lysogenic cycle, the viral DNA is incorporated into the DNA of the host, and cellular replication occurs normally. The host cell is not destroyed, and the virus does not produce progeny. A summary of the genome is depicted in Fig. 1.3. The goal of the experiment of Lwoff et al. was to understand the genesis and predictability of the lytic cycle. In their work, a collection of *E. coli* cells infected with the λ phage was irradiated by a moderate dose of ultraviolet light. In contrast to the expected scenario observed without radiation, in which many generations of cells underwent uninterrupted reproduction, the irradiated cells stopped growing and 90 minutes later a large fraction of the cells burst into a crop of λ viruses. Repetition of the experiment led to reproducibility of the result, and the phenomenon was understood to be an example of the switching on or off of a gene controlling the lytic or lysogenic fate of the cell.

Figure 1.3: Some basic features of the λ phage genome.



The next major step was the discovery of the O_R region [Ptashne \[1992\]](#). The 48502 base pairs of the genome had been sequenced, and experiments showed that between basepairs 37940 to 38020 there existed a region, the O_R , between the *ci* and *Cro* genes where the *ci* and *Cro* proteins compete on binding sites for their transcription. This region overlaps with two RNA polymerase promoters, P_{RM} , P_R , for promoting *ci* and *Cro*, respectively. P_{RM} overlaps with O_{R3} and O_{R2} , while P_R overlaps with O_{R1} , Fig. 1.4. This has given the *ci* protein the name λ repressor, since in lysogeny the *ci* protein actively represses the expression of the *Cro* gene by blocking the P_R binding site from RNA polymerase. When *Cro* is present the

affinity is highest to the P_{RM} site, blocking cI promotion.

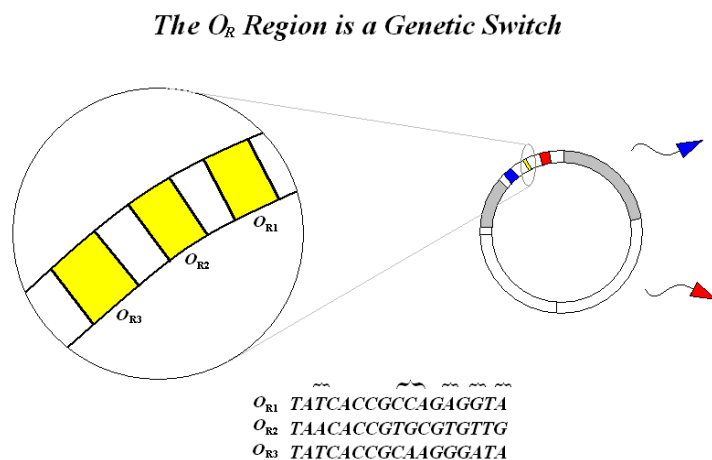


Figure 1.4: **Zooming into the O_R site reveals more structure, the P_R site overlaps with O_{R1} and O_{R2} and P_{RM} overlaps with O_{R3} .**

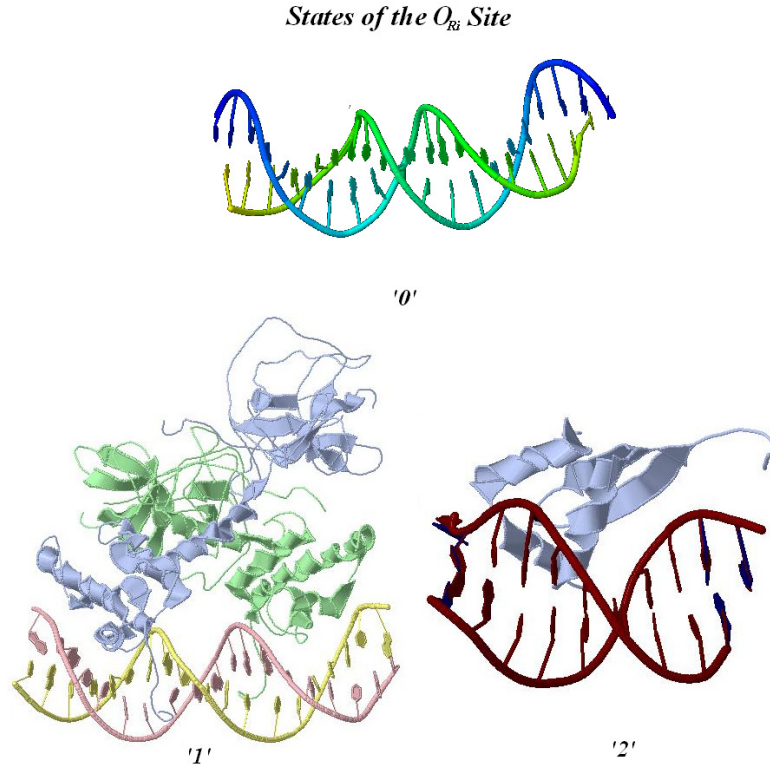
In this study, we focus on the fate of the *E. coli* bacterial cell at the moment that infection with the λ phage occurs.

In the chemistry section of this work, we introduce the equilibrium assumptions used in the model and tabulate the rates involved in some of the system kinetics, namely the binding of dimers to DNA, and the dimerization of the proteins in the cell. The ΔG values are the result of fits to experimental studies and are largely the focus of experimental work on which this model depends. All numbers have been chosen to correspond with the numbers implicit to Wang et al. [2010] which we repeat here. With modern crystallographic studies of these interactions and the state of empirical force fields, these values could in principle begin to be given a basis without the biological experiments. This avenue will not be pursued here, although it is representative of the type of hurdle involved in getting clinically useful information out of molecular structure-based studies and we stress the importance of such calculations to provide some atomic and molecular justification of this phenomenology.

In the dynamics section of this work, we review the phenomenology that has been introduced. This phenomenology is interesting in its own right as a method for exploring non-equilibrium statistical mechanics behavior over large time scales given some understanding of the equilibrium picture. The success with the λ phage is suggestive of rather general and powerful methods for exploring other gene regulatory circuits and complex systems as well. The essential content of the dynamical system of Shea and Ackers [1985] is contained in the rate equation that follows

$$\dot{x}_i = v_{x_i} \frac{Z_{x_i}(x_1, \dots, x_n)}{Z(x_1, \dots, x_n)} - \frac{x_i}{\tau_{x_i}}. \quad (1.3)$$

Figure 1.5: Depicted are three possible states of the O_{Ri} site. Each of the three DNA sites can be either empty, occupied by a cI dimer, or occupied by a Cro dimer.



Here, the time derivative of the x_i protein is written in terms of the rate, ν_{x_i} , at which RNAP and ribosome can actively create more x_i . Z_{x_i} is itself the sum over states that allow a creation process and is weighted by a sum over all states. In the cell the protein decays at a rate of τ_{x_i} . No summation convention is employed with this equation.

The inevitable limitation of Eq. (1.3) is the presence of noise, and corrections are necessary. To this end, the diffusion section of this work introduces an integral method for solving a class of partial differential equations including the Fokker-Planck equation $\partial_t P = -\partial_i(F_i P - D_{ij}\partial_j P)$. This is to provide a numerical method of solution to the general formalism for using diffusion equations to describe noise in chemical equations, Wang et al. [2008]. This is an analogy to stochastic methods for chemical equations Ge and Qian [2010], Qian et al. [2002], and path integral methods for chemical equations Wang et al. [2010]. The class is then restricted to the case of flows induced by the dynamic assumptions of the phage. This method is typical of a finite element method and its origins even predate the existence of the computer that has made it so practical Logg et al. [2012]. The following two equations are the main result of this thesis.

$$a = \int_{\Omega} (v u_{t+\delta t} + \frac{\delta t}{2} \partial_i v [D_{ij} \partial_j u_{t+\delta t} - F_i u_{t+\delta t}]) dx, \quad (1.4)$$

$$L = \int_{\Omega} (vu_t - \frac{\delta t}{2} \partial_i v [D_{ij} \partial_j u_t - F_i u_t]) dx. \quad (1.5)$$

The solutions u_t and $J = D_{ij} \partial_j u_t - F_i u_t$ are the probabilistic and flux landscapes, respectively, that appear in the title of this work.

$a = L$ is a linear matrix equation in a function space when one specifies an initial condition. The dimension of the space depends on the computational resources at our disposal and the flexibility in this direction is just one luxury of FEniCS software.⁶ The expression a is bilinear, in $u_{t+\delta t}$ and in v , when used for propagating time in the positive direction. The right hand side in this equation is only linear, depends only on v , i.e. u_t is fixed by the previous time step and thus by the initial condition.

This is done by casting all the functions into a function space, or vector function space as needed. When these expressions hold for all v in the function space, u_{t+dt} is a solution to the related Fokker-Planck equation if u_t is. Integral n th passage time equations are similarly introduced and first passage time equations are solved for all the mutants. The integral equation corresponding to $F_i \partial_i t_1(x) - D \partial_i \partial_i t_1(x) = -1$ follows

$$\int_{\Omega} (t(x)(F_i \partial_i t_1(x)) - D \partial_i t(x) \partial_i t_1(x)) dx = - \int_{\Omega} t(x) dx, \quad (1.6)$$

$$\int_{\Omega} (t(x)(F_i \partial_i t_n(x)) - D \partial_i t(x) \partial_i t_n(x)) dx = -n \int_{\Omega} t(x) T_{n-1} dx. \quad (1.7)$$

The passage times proceed recursively. After the first passage times are found and inserted for T_0 the second equation is rendered linear and the higher moments follow similarly. The situation in the O_R requires only the first passage time; the extreme value of these numbers compared to the cell cycle is how the robustness is quantified here. There is some dispersion in the literature as to what that value is, as it is difficult to measure directly, but an estimate of the upper bound is that 1 in $10^{7 \pm 2}$ flip in a cell cycle, see [Little et al. \[1999\]](#).

The integral equations are discussed in the *diffusion* section and further tabulation of the results occurs in chapter 6. The methods were kept general and modular to quickly facilitate deeper exploration as more computational resources are employed or parameters changed. It is taken for granted that the models of systems that evolve should likewise undergo evolution.

The λ phage also meets the criterion for a selective system to be worthy of intensive study; the λ phage is a modular system, a system of interacting components that can be systematically permuted [Rothstein \[1982\]](#).

He also motivates purely computational approaches to theoretical biology.

Progress in theoretical biology, we believe, needs help from

⁶Various mesh sizes were used in this thesis, a 600 by 600 space of quadratic elements, corresponding to a vector space of ≈ 720000 dimensions was about the limit of a desktop computer.

more than physics and biology. It needs help from computer science, where any [selective system] that can be described by either a statistical, mathematical, or logical theory can be modeled. In particular, reconfigurable systems consisting of interacting subsystems should be studied intensively.â€Ž

The λ phage has proven to be a case in point of the benefits of such intensive study. The effects of some mutations in the O_R have already been experimentally realized, [Little et al. \[1999\]](#), and λ_{121} , λ_{323} and $\lambda_{323'}$ have been experimentally synthesised and studied. The most successful exploration in this varietal has been by far directly from the biologists, where DNA and proteins are directly manipulated and observed. It is the authors hope that in time the contributions from physics, mathematics, computer science, and chemistry will provide the means to observe much more quickly and cost effectively some of the effects possible in this vast configuration space, and direct biological experiments towards interesting possibilities. It is largely in the direction of this ideal that this thesis explores, and we will simulate those and similar mutations, proposing theoretical predictions to feasible experiments which could further constrain the models.

1.4 Evolution

Evolution, like gravity and thermodynamics, is so seamlessly entangled into our lives as to often go unnoticed. Once, however, the concept is understood it is quite self-evident, and is as readily seen as the selection of an item off of a menu or the selection of students or employees from a pool of applicants. This interplay between chance and determinism, the arbitrary and the necessary, and populations and their individuals provides more than enough room for complexity to emerge, even before granting a capacity for volition to the individuals. Much confusion about evolution can be avoided by clearly separating what one is selecting from and selecting for, [Mayr \[2002\]](#). This thesis takes leave from the *noblesse oblige*, in speaking of matters in which one is not an expert,⁷ but the motivation of such an investigation as this is largely rooted in Evolution, and such a discovery warrants discussion in a physics thesis. Indeed in [Ao \[2008\]](#) it is speculated that Evolution *is* the theory of everything physicists have vainly sought after. In addition, the main methodology of this thesis is one of tinkering; to do lots of trials and make a lot of errors. This process is in correspondence with how natural selection naturally selects, see [Jacob \[1977\]](#).

1.4.1 Natural Selection from...

We will be considering natural selection from the set of DNA configurations. We choose 27 of these configurations corresponding to mutations in the O_R region

⁷somebody has to bridge the gaps...[Schrödinger \[1992\]](#)

for which present knowledge enables a prediction of what the binding affinities would be. It is demonstrated that just this restricted space is enough to sample both sides of a bifurcation.

1.4.2 *Natural Selection for...*

If biological systems are like other physical systems they seek to minimize energy and maximize entropy. We will rank our mutants from fittest to least fit based on passage times, energy, and entropy. Natural selection is a fine balancing of many considerations.

1.5 *Three modes of Computation*

Many modes of computing with molecules have been proposed and these can be split into three overlapping categories.

Although the number of genes in the human genome has likely been constant for a few million years, the number in the refereed literature steadily decreases with time. In the human genome there are believed to be around 25000 protein coding genes, and much more of the genome is understood to be dedicated to regulation and signaling processes than initially thought.

Cooperativity between regulatory proteins and RNAP for transcription is seeming to be the rule rather than the exception [Ptashne \[2011\]](#). The signaling involved masterfully involves the time domain and is the major port of departure between DNA type computation and a Turing tape. Whereas a Turing machine requires a head to physically transform the bit strings along the tape, DNA strictly avoids this phenomenon and instead prefers to modify the tape through binding of ligands. This is the tragic flaw of Benett's proposed DNA computer in 1982 in which a ribosome like structure alters base pairs. It may be more easily implemented on RNA than DNA, but still the lesson to be learned is to program with the time domain as much as possible. The paradigms of causality are being replaced by the paradigms of synchronicity; GPU, DuoCore, QuadCore, quantum computation, and chemical computation are all parallel computations that rely on a balancing of rates in order for their components to sync and perform properly.

1.5.1 *Binding Sites*

Anything that can count can compute, and RNAP transcribing a gene into a protein is such a means of computing. This is the type of computation discussed quantitatively in this thesis, using the phage λ as a concrete example. In the dynamics section we will discuss the rich space of analog computation, which computation with binding sites can simulate.

1.5.2 Polymerization

A type of computation DNA can employ, in which a large configuration space is rapidly sampled through polymerization processes has demonstrated working, fast algorithm for NP problems. [Adleman \[1994, 1998\]](#), [Braich et al. \[2002\]](#)

1.5.3 Solitons

A method of computation that macromolecules could employ is through solitons. The effective field theories for some of the empirical force fields used to describe the molecular dynamics are known to emit soliton solutions, [Yakushevich et al. \[2002\]](#). The presence of soliton solutions is sufficient to define a bit, and the propagation of the soliton is a sufficient means to transfer a signal. The empirical model used in protein folding is similar in structure to the empirical models used in superconducting circuits. Cascading logic implemented through solitons in superconducting circuits is the most promising approach to superconducting-based computation, [Likharev and Semenov \[1991\]](#), [Semenov et al. \[2003\]](#), [Ren et al. \[2009\]](#), [Ren \[2011\]](#). It will be found that biological systems have the same physical limitations as engines or other computational devices and deep understanding of those limitations are required, [Likharev \[1982\]](#).

To specify, consider the Hamiltonian used in [Yakushevich et al. \[2002\]](#),

$$H = \sum_n \left(\frac{1}{2} I_{n,1} \dot{\phi}_{n,1}^2 + \frac{1}{2} I_{n,2} \dot{\phi}_{n,2}^2 + \epsilon_{n,1} \sin^2 \left(\frac{\phi_{n+1,1} - \phi_{n,1}}{2} \right) + \epsilon_{n,2} \sin^2 \left(\frac{\phi_{n+1,1} - \phi_{n,2}}{2} \right) + V_{\alpha\beta}(\phi_{n,1}, \phi_{n,2}) \right). \quad (1.8)$$

Now consider a Hamiltonian used in superconducting circuits [Averin et al. \[2006\]](#),

$$H = \sum_n \left[\frac{Q_n^2}{2C} + 2E_J \sin^2 \left(\frac{\phi_n}{2} \right) + E_L (\phi_{n+1} - \phi_n - \phi_n^{(e)})^2 \right]. \quad (1.9)$$

We can see that the correspondence is not just a matter of romance. Letting the nucleotides correspond to junctions, and the phase of the superconducting order parameter correspond to physical angles, one can proceed quite formally.

The solitons would then be of the sine-Gordon variational and approachable through standard methods, [Rajaraman \[1987\]](#).

It is a mystery of gene expression and cell behaviour how genes turn on and off, how several distinct and large regions of DNA, genes, can be quiet for the majority of a cell cycle and then be triggered for expression of certain proteins at critical moments. Although these processes are often known to be caused by chemical triggers, mechanical triggers are not ruled out, and anything that can induce long-range spatial and temporal correlations over a chain of DNA should

be considered deeply. Solitons typify long-range spatial coordination, and are at least an entertaining possibility of signaling modes available in DNA.

The analogy is suggestive also of modeling systems via direct mimicry. Junctions could be chosen with capacitance and junction energies to directly simulate on a transmission line the dynamics of a DNA chain. Similar circuits can be designed for protein folding with a correspondence between the junctions and the amino acids. Although DNA and proteins are brilliant, they are not particularly fast with that brilliance, and such circuits could boast protein folding faster than proteins can fold a protein. The lesson to be learned in superconducting circuits from DNA is a promise of diverse effects emerging from the aperiodicity of a crystal. One might be capable of developing dramatically sophisticated circuits, without a large number of components, by using similar-but-different junctions. The richness of the protein functions themselves are believed to be derivative of their structure, which itself is yet another chain, this time of geometric motifs, helices, hairpins. One is left curious to find analogues of these structures in the circuits.

1.6 *Methods of Numerical Analysis*

The rich diversity intrinsic to biological systems lends necessity towards mathematics and computation that is equally rich and diverse. In computational terms, biophysics necessitates modularity, as code needs to be useful and applicable in a wide variety of settings. Emphasis has been placed on exactly this point throughout the design of numerical investigations reported in this thesis. This outlook has allowed for scanning large regimes of parameters of particular systems as well as investigating many systems and models. It is a well known problem of computer science that generality comes at the cost of requiring large amounts of code to consider all possible cases. This problem has largely been circumvented through the use of automated form compilation, in which the code written is designed to write more specific code when a particular instance is requested, [Logg et al. \[2012\]](#)⁸. Such a method lends to an elegant correspondence between the mathematics describing a problem and the corresponding computer instruction.

The mathematics employed for this analysis is a modern incarnation of finite element methods developed before computers existed to provide a general method to solve or approximate large classes of partial differential equations from theories of hydrodynamics, electromagnetism, and elasticity, and is sufficient to solve the partial differential equations arising from biophysics as well. We find that even in circumstances where analytic solutions exist and can be found, numerics are more illuminating than the resulting expression and provide greater accessibility to the particular values of a function at given points in the domain. An analytic

⁸One cannot overstate the importance of the software and contributors therein. Their achievement is momentous.

expression may require just as much numerical work to, say, evaluate and sum the necessary hypergeometric functions when a particular value at a particular point for a particular set of parameters. Thus, even in instances where symbolic analysis of solutions are accessible we insist on numerics.

Another important necessity of computational techniques used to study biological systems is scalability. Computer instructions, if not the computer itself, should easily adapt as numbers of parameters, degrees of freedom, and resolution of meshes increase. Equally as diverse as the biological systems themselves are the assertions found in the literature about the most relevant degrees of the system, the rate of changes exhibited by the system, the parameter values the system exhibits, etc. Scalability ensures that the choices made in approximating a system can be based on physics or other considerations and not depend exclusively on access to computational power. For example, better computers should result in better approximations but not be a prerequisite for the utility of the method; indeed, the use of these methods predates the existence of the computer. The desire for scalability also motivated the use of FEniCS software which can be parallelized for larger systems and meshes. The scaling that these methods utilize are suboptimal in that as the number of system degrees of freedom are increased, the computational resources required for the same resolution increases exponentially. For an optiplex duo core, a computer of modest price and performance, we found this to limit the size of mesh to two dimensions with 600 quadratic Lagrange elements in each direction. Three-dimensional queries required the use of adaptive mesh techniques to locate the active regions of interest and localize the resolution to these areas. The computer instructions, the human interface to the computer, scale ideally. That is to say, the only instructions to change as the systems analyzed change are the definitions of the system.

This partial scaling manifests in the code simply crashing when more resources are requested than available, but would continue to work on machines or clusters where the resources are indeed available. A pleasant corollary of the active development of FEniCS is that code already written ages like a fine wine. The benchmarks stated have changed over time, constantly speeding up, and one only needs patience for the physical interface to be as fast as one desires, as the human interface is nearly optimized. If you want a solution, it is now as easy as a well-defined question.

Another important aspect of the computation, to the author, is that all results and methods are reproducible and extendable from open-source utilities. This facilitates the transfer of knowledge from the more fortunate to the less fortunate and simply evolves faster than closed source analogues. That is not to say that open-source is better than closed-source so much as that it will be better, and contributions in the direction of automated form compilation are quickly eliminating the gap between the quality of freely available symbolic mathematics code and the privatized software for the same tasks.

1.6.1 FEniCS

The diffusion equations herein would never have been solved without the expertise and their freely available software. If the codes used throughout this thesis are to be used successfully a basic understanding of not only object-oriented programming is necessary, but also understanding and installation of the software available at www.fenicsproject.org is necessary.

Data aequatione, spoketh Newton, quotcunque fluentes quantitae involvente fluxiones invenire et vice versa. Which Arnold put into modern mathematical prose: *It is useful to solve differential equations.* Finite elements do just that.

1.6.2 ROOT

ROOT is an open source utility developed at CERN and the standard analysis framework used at all major particle accelerators of the world, [CER \[2011\]](#). The object-oriented tools like graphing and histogramming utilities largely extend the utility and accessibility of data generated in this thesis. The tools long since built for analysing nuclear data are readily available for the analysis of other data as well and provide automatable procedures for problems such as fitting peaks and simple accounting.

In particular, a pyROOT paradigm was embraced to combine this functionality with the ease of python and integrate seamlessly with the python tools built by FEniCS.

1.6.3 Aleph

It is taken for granted that models of systems that evolve should evolve as well, and Launchpad documents the code and the evolution of this code, [Borggren \[2011\]](#).

1.7 Overview

This work is largely a survey; it is the study of a biological system with the techniques of applied mathematics, engineering, computer science, and chemistry, directed at an audience of physicists and philosophers. As such it runs the risk of being neither here nor there, to shoot three bears while only infuriating them. To this end offense is inevitable. An ISI Web of Knowledge query reveals 170,303 results for Finite Elements, 11,445 for λ Phage, and 3,151,834 for genes. To contrast, Quantum Mechanics appears 31,673 times and Quantum Hall Effect 6,557 times. Thus, the references are largely a matter of taste, and do not attempt a reflection of history, so much as what I felt to be instructive accounts of the systems and useful ports of entry to the historical literature.

A goal is maintained to provide a perspective on generality while keeping the λ phage as a concrete example. This work seeks to explain to a physical audience the details of a biological system.

The philosophy of physics and information took inspiration from many places: [Anderson \[1972, 2000\]](#), [Bennett \[1982\]](#), [Toffoli \[1982\]](#), [Likharev \[1996, 1982\]](#), [Landauer \[1961\]](#), and [Rothstein \[1982\]](#).

In matters of solitons there is [Yakushevich et al. \[2002\]](#), [Averin et al. \[2006\]](#), [Likharev and Semenov \[1991\]](#), [Likharev \[1996\]](#).

For background and data in the λ phage, see [Ptashne \[2011, 1992\]](#), [Ptashne and Gann \[2002\]](#), [Little et al. \[1999\]](#), [Atsumi and Little \[2006\]](#).

For evolution there are good introductions, see [Mayr \[2002\]](#), [Jacob \[1977\]](#).

For mathematics, [Arnold \[1978\]](#), [Perko \[2001\]](#).

For early modeling and the head node for the taxonomy tree of the ΔG values, [Ackers et al. \[1982\]](#), [Shea and Ackers \[1985\]](#).

Later modeling, [Aurell and Sneppen \[2002\]](#), [Aurell et al. \[2002\]](#), [Zhu et al. \[2004\]](#), [Wang et al. \[2010\]](#).

The methods and ideas herein are largely in correspondence to: [Wang et al. \[2008\]](#), [Zhu et al. \[2004\]](#), [Wang et al. \[2008, 2010\]](#), [Ge and Qian \[2010\]](#), [Feng and Wang \[2011\]](#) and have benefited from numerous conversations with Jin Wang.

An accomplishment of this thesis is one of method; it is the computational framework in [Borggren \[2011\]](#), this is but a modest interface and synthesis of tools [Logg et al. \[2012\]](#), [CER \[2011\]](#).

2 Binding Polynomials and the Chemistry of the O_R Region

*Pray be under no constraint in this house. This is Liberty-hall, gentlemen.
You may do just as you please here.
Oliver Goldsmith, She Stoops to Conquer*

2.1 ΔG Measurements

In this section we consider the partition functions used to describe the chemical binding interactions in the O_R region that must be posited before we proceed to introduce dynamics. There is some choice so we explore multiple possibilities.

For one, the original partition functions include explicitly RNA polymerase at a fixed concentration, whereas many new pictures also include bindings to the O_L region. A complication is that both cI and Cro proteins bind to any of the 3 sites, on O_R or O_L in their dimer form. It is also now established that up to octamers can be present and, amazingly, cI octamers can bind at once to the O_R as the O_L and short the DNA loop exactly across the cI gene. We will continue to use the equilibrium relationship between dimer and monomer concentrations in our partition function, [Reinitz and Vaisnys \[1990\]](#), [Wang et al. \[2010\]](#), [Aurell and Sneppen \[2002\]](#), [Zhu et al. \[2004\]](#).

Again, x, y are cI, Cro protein numbers and $[cI], [Cro]$ denote the dimer concentrations. The simplest partition function used does not include RNAP binding affinities or concentration, and consists of the 27 states permitted by having nothing, cI, or Cro on each of the three sites. Of these states, 1 is zero liganded, 6 are singly liganded, 12 are doubly liganded, and 8 are triply liganded¹ The nature of the partition function is perhaps most evident when written in the form

¹Since ΔG values are negative and the individual processes occur at similar rates to each other, this establishes a sort of band structure; the triply liganded configurations occur the most often, then the doubly liganded, then the singly liganded, and then the scarcest event is to find an open DNA configuration.

$$\begin{aligned}
Z = 1 + & \\
& [cI] \left(e^{-\frac{\Delta G_{001}}{RT}} + e^{-\frac{\Delta G_{010}}{RT}} + e^{-\frac{\Delta G_{100}}{RT}} \right) + \\
& [Cro] \left(e^{-\frac{\Delta G_{002}}{RT}} + e^{-\frac{\Delta G_{020}}{RT}} + e^{-\frac{\Delta G_{200}}{RT}} \right) + \\
& [cI]^2 \left(e^{-\frac{\Delta G_{011}}{RT}} + e^{-\frac{\Delta G_{101}}{RT}} + e^{-\frac{\Delta G_{110}}{RT}} \right) + \\
& [Cro]^2 \left(e^{-\frac{\Delta G_{022}}{RT}} + e^{-\frac{\Delta G_{202}}{RT}} + e^{-\frac{\Delta G_{220}}{RT}} \right) + \\
& [cI][Cro] \left(e^{-\frac{\Delta G_{012}}{RT}} + e^{-\frac{\Delta G_{021}}{RT}} + e^{-\frac{\Delta G_{102}}{RT}} + \right. \\
& \left. e^{-\frac{\Delta G_{201}}{RT}} + e^{-\frac{\Delta G_{210}}{RT}} + e^{-\frac{\Delta G_{120}}{RT}} \right) + \\
& [cI]^3 \left(e^{-\frac{\Delta G_{111}}{RT}} \right) + \\
& [Cro]^3 \left(e^{-\frac{\Delta G_{222}}{RT}} \right) + \\
& [cI]^2 [Cro] \left(e^{-\frac{\Delta G_{112}}{RT}} + e^{-\frac{\Delta G_{121}}{RT}} + e^{-\frac{\Delta G_{211}}{RT}} \right) + \\
& [cI][Cro]^2 \left(e^{-\frac{\Delta G_{122}}{RT}} + e^{-\frac{\Delta G_{212}}{RT}} + e^{-\frac{\Delta G_{221}}{RT}} \right)
\end{aligned} \tag{2.1}$$

This form makes it clear that it is symmetric with respect to permutation of the sites. The argument of [Aurell and Sneppen \[2002\]](#) is that RNAP needs an open P_R site to promote Cro and an open P_{RM} site to promote cI, and if the site is open, RNAP will be overwhelmingly first in line to bind and promote such states. It is known that the P_R site overlaps with O_{R1} and O_{R2} and P_{RM} overlaps with O_{R3} . This symmetry is not present for the subspaces; the sum over the substates with a free P_{RM} can promote x, a cI monomer, and the sum over the substates with a free P_R site can promote y for a Cro monomer

$$\begin{aligned}
Z_x = 1 + & \\
& [cI] \left(e^{-\frac{\Delta G_{010}}{RT}} + e^{-\frac{\Delta G_{100}}{RT}} \right) + \\
& [Cro] \left(e^{-\frac{\Delta G_{020}}{RT}} + e^{-\frac{\Delta G_{200}}{RT}} \right) + \\
& [cI]^2 e^{-\frac{\Delta G_{110}}{RT}} + \\
& [Cro]^2 e^{-\frac{\Delta G_{220}}{RT}} + \\
& [cI][Cro] \left(e^{-\frac{\Delta G_{210}}{RT}} + e^{-\frac{\Delta G_{120}}{RT}} \right), \\
Z_y = 1 + & \\
& [cI] e^{-\frac{\Delta G_{001}}{RT}} + \\
& [Cro] e^{-\frac{\Delta G_{002}}{RT}}.
\end{aligned} \tag{2.2}$$

with

$$[cI] = [x * pnc] / 2 + e^{\Delta G_{cI} / RT} / 8 - ([x * pnc] \frac{e^{(\Delta G_{cI} / RT)}}{8} + \frac{e^{2\Delta G_{cI} / RT}}{64})^{1/2} \quad (2.3)$$

$$[Cro] = [y * pnc] / 2 + e^{\Delta G_{Cro} / RT} / 8 - ([y * pnc] \frac{e^{(\Delta G_{Cro} / RT)}}{8} + \frac{e^{2\Delta G_{Cro} / RT}}{64})^{1/2} \quad (2.4)$$

Interlude: On dimerization.²

The key feature to notice here is that, since all the binding affinities are negative, the sites are dominated by the triply liganded terms. The occurrence of an x promoting state is already rare, and a y promotion state exponentially more scarce than those. The partition function here is seen to be a sum over monomials of the forms, 1, $[cI]$, $[Cro]$, $[cI]^2$, $[Cro]^2$, $[cI][Cro]$, $[cI]^3$, $[Cro]^3$, $[cI]^2[Cro]$, and $[Cro][cI]^2$. In the original Shea and Ackers contribution we note that additional monomials of the form $[RNAP]$, $[RNAP][cI]$, $[RNAP][Cro]$, $[RNAP][cI]^2$, $[RNAP][Cro]^2$, $[RNAP][cI][Cro]$ and $[RNAP]^2$ were also included. This contributed to 13 more states included in the partition function for a total of 40. Only 40 of the 3^4 states are permissible since an RNAP at P_R precludes any binding to O_{R1} and O_{R2} . The $-\Delta G_s$

² On Dimerization

1 + 1 = 1

Una goccia piú una goccia, fanno una goccia piú grande, non due!
the Tarkovsky equation and explanation, *Nostalgia*

Todos los fuegos el fuego
Julio Cortazar, *Todos los fuegos el fuego*

The dimerization equation used in this work originates from a quasi-steady state assumption for the dimerization reaction.

$$[cI_1]^2 = [cI_2] e^{\frac{\Delta G_{cI}}{RT}}, \quad [N_{cI}] = [cI_1] + 2[cI_2] \quad (2.5)$$

This gives a quadratic equation with roots giving the cI dimer concentration, $[cI_2]$ in terms of the total protein concentration, $[N_{cI}]$.

$$[cI_2] = \frac{[N_{cI}]}{2} + \frac{e^{\frac{\Delta G_{cI}}{RT}} \pm \sqrt{e^{\frac{2\Delta G_{cI}}{RT}} + 8e^{\frac{\Delta G_{cI}}{RT}} [N_{cI}]}}{8} \quad (2.6)$$

The root with the positive sign is the expression stated in Wang et al. [2010], Ao [2004] and the negative sign corresponds to the expression used in the calculations of Reinitz and Vaisnys [1990], Ao [2004], Wang et al. [2010]. The negative sign seems to be the better physical choice in that, in the limit of zero monomer concentration, one also has zero dimer concentration.

λ_{123} , wild – type O_R affinities (kcal/mol)					ΔG_{ijk}		
protein	O_{R1}	O_{R2}	O_{R3}	ref.	equation	monomial	premotors
$\Delta G_{001}, \Delta G_{010}, \Delta G_{100}$	-15	-13	-12	Zhu et al. [2004]	$\Delta G_{000} = 0.0$	1	Z_x, Z_y
$\Delta G_{002}, \Delta G_{020}, \Delta G_{200}$	-18.4	-17.1	-19.5	Zhu et al. [2004]	$\Delta G_{001} = -12.0$	$[cI]$	Z_y
$\Delta G_{coop} = -6.9$ $pnc = 1.5 * 10^{-11}$ $mpi = 3^a$					$\Delta G_{010} = -13.0$	$[cI]$	Z_x
					$\Delta G_{100} = -15.0$	$[cI]$	Z_x
					$\Delta G_{002} = -19.5$	$[Cro]$	Z_y
					$\Delta G_{020} = -17.1$	$[Cro]$	Z_x
					$\Delta G_{200} = -18.4$	$[Cro]$	Z_x
$\Delta G_{ijk}(cont)$		$\Delta G_{011} = \Delta G_{010} + \Delta G_{001} + \Delta G_{coop}$			$[cI]^2$		
		$\Delta G_{110} = \Delta G_{010} + \Delta G_{100} + \Delta G_{coop}$			$[cI]^2$		Z_x
		$\Delta G_{101} = \Delta G_{100} + \Delta G_{001}$			$[cI]^2$		
		$\Delta G_{022} = \Delta G_{020} + \Delta G_{002}$			$[Cro]^2$		
		$\Delta G_{220} = \Delta G_{020} + \Delta G_{200}$			$[Cro]^2$		Z_x
		$\Delta G_{202} = \Delta G_{200} + \Delta G_{002}$			$[Cro]^2$		
		$\Delta G_{120} = \Delta G_{100} + \Delta G_{020}$			$[cI][Cro]$		Z_x
		$\Delta G_{210} = \Delta G_{200} + \Delta G_{010}$			$[cI][Cro]$		Z_x
		$\Delta G_{102} = \Delta G_{100} + \Delta G_{002}$			$[cI][Cro]$		
		$\Delta G_{201} = \Delta G_{200} + \Delta G_{001}$			$[cI][Cro]$		
		$\Delta G_{012} = \Delta G_{010} + \Delta G_{002}$			$[cI][Cro]$		
		$\Delta G_{021} = \Delta G_{020} + \Delta G_{001}$			$[cI][Cro]$		
		$\Delta G_{111} = \Delta G_{100} + \Delta G_{010} + \Delta G_{001} + \Delta G_{coop}$			$[cI]^3$		
		$\Delta G_{222} = \Delta G_{200} + \Delta G_{020} + \Delta G_{002}$			$[Cro]^3$		
		$\Delta G_{112} = \Delta G_{100} + \Delta G_{010} + \Delta G_{coop} + \Delta G_{002}$			$[cI]^2[Cro]$		
		$\Delta G_{121} = \Delta G_{100} + \Delta G_{020} + \Delta G_{001}$			$[cI]^2[Cro]$		
		$\Delta G_{211} = \Delta G_{200} + \Delta G_{001} + \Delta G_{010} + \Delta G_{coop}$			$[cI]^2[Cro]$		
		$\Delta G_{221} = \Delta G_{200} + \Delta G_{020} + \Delta G_{001}$			$[cI][Cro]^2$		
		$\Delta G_{212} = \Delta G_{200} + \Delta G_{010} + \Delta G_{002}$			$[cI][Cro]^2$		
		$\Delta G_{122} = \Delta G_{100} + \Delta G_{020} + \Delta G_{002}$			$[cI][Cro]^2$		

Table 2.1: Values of variables referenced by the partition function.

^ampi is phages per cell.

values for these states are also actively measured by the experimental community and could be included in the partition function insofar as RNAP does indeed bind in the O_R , so that in any given interval of time there is indeed the probability that the O_R is in a state not considered in the above partition function. The effect is said that eliminating the RNAP is *constant*, though there is a continuation in the literature that prefers to keep RNA polymerase explicitly in the partition function and no mention to what the value of that *constant*. The situation is a little subtler then that and the value of this effect, as shown here, relates to the RNAP binding energies to the P_R , and P_{RM} as well as absolute concentrations. The benefit is still a simpler partition function and since RNAP is fixed for the simulation, the additional monomials do not alter the algebraic structure of the equations but only append to the weights of the already present monomials, but this effect can still

induce bifurcations and one should be careful.

It is then necessary to write the dimer concentrations in terms of the monomer number. Here we use the expression implicit to the results of [Zhu et al. \[2004\]](#), [Wang et al. \[2010\]](#), where we explicitly show how the number pnc , labeled as a conversion factor from number to concentration, enters into their equation.

We note that the above equations appear with the mistake of a positive sign, in [Wang et al. \[2010\]](#) and the arxiv version of [Zhu et al. \[2004\]](#), and use the expression from [Reinitz and Vaisnys \[1990\]](#) with the insertion of the pnc , a converting factor from protein number to concentration, and implicit assumptions about RNAP (see the note on dimerization). The bacterial volume used in [Zhu et al. \[2004\]](#) is stated to be $.7 * 10^{-15}L$ and pnc is in inverse proportion to the volume. The value used for pnc is $1.5 * 10^{-11}$. A calculation reveals and is corroborated by a private communication that in actuality $pnc = (volume * N_A)^{-1} = 2.37 * 10^{-9} Moles/L$ and the listed value is a typo. Use of that number reveals unrealistic protein concentrations and it is this authors belief that the number pnc includes implicit assumptions about the RNAP concentrations and affinities. Provided RNAP is removed from the partition function, a pnc value of $1.5 * 10^{-11}$ is indeed more consistent with the numbers obtained from experiment, and was likely tuned ad hoc to the present value for that reason. No reference is made about the value of the number or conversion in [Wang et al. \[2010\]](#), although the calculations present suggest the use of the conversion from Ao with the equation from Reinitz despite the errors in their written equation. Many calculations should be performed twice, once with the values similar to [Zhu et al. \[2004\]](#) and again with RNAP in the partition function and $pnc = 2.37 * 10^{-9}$ which is postponed for future work. As noted in the introduction this equation should be modified in future work anyway as a result of the higher order polymerization processes that occur.

2.2 A Partition Function; RNAP or no RNAP?

The partition function in [Shea and Ackers \[1985\]](#) reads, with $[RNAP]$ the concentration and assigning the index '3' to denote the RNAP occupancy.

$$\begin{aligned}
Z = & 1 + ([RNAP](e^{-\Delta G_{03}} + e^{-\Delta G_{30}}) + [RNAP]^2 e^{-\Delta G_{33}}) + \\
& [cI](e^{-\Delta G_{001}} + e^{-\Delta G_{010}} + e^{-\Delta G_{100}} + [RNAP](e^{-\Delta G_{31}} + e^{-\Delta G_{103}} + e^{-\Delta G_{130}})) + \\
& [Cro](e^{-\Delta G_{002}} + e^{-\Delta G_{020}} + e^{-\Delta G_{200}} + [RNAP](e^{-\Delta G_{32}} + e^{-\Delta G_{203}} + e^{-\Delta G_{230}})) + \\
& [cI]^2(e^{-\Delta G_{011}} + e^{-\Delta G_{101}} + e^{-\Delta G_{110}} + [RNAP]e^{-\Delta G_{113}}) + \\
& [Cro]^2(e^{-\Delta G_{022}} + e^{-\Delta G_{202}} + e^{-\Delta G_{220}} + [RNAP]e^{-\Delta G_{223}}) + \\
& [cI][Cro](e^{-\Delta G_{012}} + e^{-\Delta G_{021}} + e^{-\Delta G_{102}} + e^{-\Delta G_{201}} + e^{-\Delta G_{210}} + e^{-\Delta G_{120}} + \\
& + [RNAP](e^{-\Delta G_{123}} + e^{-\Delta G_{213}})) + \\
& [cI]^3(e^{-\Delta G_{111}}) + \\
& [Cro]^3(e^{-\Delta G_{222}}) + \\
& [cI]^2[Cro](e^{-\Delta G_{112}} + e^{-\Delta G_{121}} + e^{-\Delta G_{211}}) + \\
& [cI][Cro]^2(e^{-\Delta G_{122}} + e^{-\Delta G_{212}} + e^{-\Delta G_{221}})
\end{aligned} \tag{2.7}$$

It is clear that the effect of $[RNAP]$ is not constant; $[RNAP]$ does not factor out of the partition function. Three experimental numbers, the $RNAP$ concentration, the $RNAP$ binding to the cI promoter, and an $RNAP$ binding to the Cro promoter are entangled into this partition function. Removing RNA polymerase from the partition function also has the effect of disturbing the time scales, in that equation 2.1 only accounts for the fraction of a dt in which no RNA polymerase is bound.

We note that this form of the partition function would require corrections for low numbers in that it uses the approximations $n^2 \approx n * (n - 1)$, $n^3 \approx n(n - 1)(n - 2)$. It is likely that some of the complexity of these systems near the axis is residual of such an approximation. This becomes especially more important if multiple phages have invaded the cell or the O_L is included in the partition function. Implicit to the figures in Zhu is the T_r , T_{rm} , and T_{rr} values reported are 1/3 of the values used in the calculation. This is due to the inclusion of 3 phages in the cell as in Aurell and Sneppen [2002]. The calculation left out presumes that $Z' = Z^3$, and the states that promote a ci , for example, are enhanced by the number of phages present $\frac{3 * Z_x * Z^2}{Z^3}$. Also, it is often presumed that for the composite system of the left and right operator regions $Z' = Z_{OL} Z_{OR}$. For a partition function to be the product of the subsystems, the subsystems must not interact. This is not the case here, and especially for the low numbers involved. One need only plot, for example, x^9 next to $x * (x - 1) * (x - 2) * (x - 3) * (x - 4) * (x - 5) * (x - 6) * (x - 7) * (x - 8)$ to see that the difference is considerable. This error is further increased with monomers present as well, since this too reduces the number of dimers and increases the fluctuation.

We use $\Delta G_{cI} = -11.1$, $\Delta G_{Cro} = -7.0$, and $\Delta G_{coop} = -6.9$. The states and their values are collected in Table 2.1

The values for the multiply liganded states are found by summing up the values of the constituent singly liganded states. Cooperativity gives an additional sum, $\Delta G_{coop} = -6.9$, for the states with cI and cI on adjacent sites. These are the G_{*11} , G_{1*1} configurations where $*$ can denote 0,1, or 2. $RT = .617 kcal/mol$ is the Boltzmann factor for the 310 Kelvin laboratory conditions which these were accumulated. When we speak of the mutant λ_{ijk} we mean that the affinity for the i th site is used in place of O_{R_1} , j th in place of O_{R_2} , and k th in place of O_{R_3} . Some of these are likely biologically unstable but it is inevitable that different strings for the operator configurations may have similar ΔG and thus similar dynamics.

3 Dynamics Near Equilibrium

The key to the treasure is the treasure.
John Barth, *Chimera*

3.1 Gene Expression and General Dynamical Systems

Although this is a simplified model, the need for general dynamical systems is already evident.

The number of stationary points is complicated by dimerization but in the basic case this number is found to be in proportion to the $(\text{number of binding sites} + 1) \times (\text{number of protein species})$. The positivity of the partition function and concentration ensures that some of these are physically realizable. The dimerization used here doubles the number of stationary points. Thus it is reasonable to start from very general mathematical considerations [Liapounoff \[1893\]](#), [Perko \[2001\]](#) and well over a century of elegant mathematics that would be required to understand the mathematics of gene expression. One is compelled to introduce a modified model that is more general and symmetric in the form in which creation and annihilation processes enter. For example, consider the n equations to replace for the discussion equation 1.3.

$$\dot{x}_i = \frac{\nu_{x_i} Z_{x_i}(x_1, \dots, x_n; H_2O, RNAP, Ribosomes, \dots) - \nu_{x_i^+} Z_{x_i^+}(x_1, \dots, x_n; H_2O, RNAP, Ribosomes, \dots)}{Z(x_1, \dots, x_n; H_2O, RNAP, Ribosomes, \dots)} \quad (3.1)$$

Here one can begin with a partition function for the n proteins and let quantities like H_2O , RNAP, all the things that contribute non-negligible partial pressures in the cell, enter as control parameters. Say, $Z(x_1, \dots, x_n; H_2O, RNAP, Ribosomes, \dots)$. Again ν_{x_i} is the rate at which transcription in the Z_{x_i} states occur. Introduced is an expression for annihilation, $\nu_{x_i^+} Z_{x_i^+}$, to account for the specific events that destroy a protein. For example, in the λ phage situation, the cI protein is destroyed by another protein, RecA, which destroys a cI monomer by cutting it in half. In such a situation one would include a term $Z_{cI^+} \propto [cI][RecA]$ governed by a rate, ν_{cI^+} of this destruction process.

The motivation is that the stationary points would appear at the points x_1, \dots, x_n where $\nu_{x_i} Z_{x_i}(x_1, \dots, x_n; H_2O, RNAP, Ribosomes, \dots) - \nu_{x_i^+} Z_{x_i^+}(x_1, \dots, x_n; H_2O, RNAP, Ribosomes, \dots) =$

0 for all i . This is a set of polynomial equations in the protein variables x_i . This allows one to either fit to known systems, or construct others. Also since, $Z_{x,y,z}$ is positive, and any term appearing in the numerator appears with a positive sign in the denominator, many qualitative features persist. Say $\dot{x} = f(x)$ and $\dot{y} = f(y)/g(y)$. One can see that if $\dot{x} = 0$, then $\dot{y} = 0$, since $g(y)$ is non-negative that the stationary points in the first system are the same as in the latter. Next, in the neighborhood of a stationary point the dynamics are qualitatively the same as the linear system,

$$\dot{x} \approx \left. \frac{\partial f}{\partial x} \right|_{x=x_{\text{stationary}}} (x - x_{\text{stationary}}), \quad (3.2)$$

and

$$\dot{y} \approx \left. \frac{\frac{\partial f}{\partial y} g(y) - \frac{\partial g}{\partial y} f(y)}{g(y)^2} \right|_{y=y_{\text{stationary}}} (y - y_{\text{stationary}}). \quad (3.3)$$

Since at a stationary point $f(y) = 0$, evidently,

$$\dot{y} \approx \left. \frac{\partial f}{\partial y} \right|_{y=y_{\text{stationary}}} (y - y_{\text{stationary}}). \quad (3.4)$$

The two systems have similar quantitative behavior then too, with the dynamics simply scaled:

$$t \rightsquigarrow t * g(y) \Big|_{y=y_{\text{stationary}}}. \quad (3.5)$$

For an overview of such methods in dynamical systems many texts are available, [Perko \[2001\]](#).

This interlude is to motivate some of the extent to which these kinetics serve as analog circuits. Restricting the numerators to polynomials, $f_i(x_1, \dots, x_n)$, in which f_i is non-negative when $x_i = 0$, that is to say we require proteins to exist before we can destroy them. Then we can always *program a set of binding-expression interactions* to simulate such dynamics.

For example, say one would like to construct a circuit with dynamics similar to $\dot{x} = x + x^2 - x * z$, $\dot{y} = y + y^2 - x * y$, $\dot{z} = z + z^2 - y * z$. Weighting by $Z_{x,y,z} = 1 + x + x^2 + y + y^2 + z + z^2 + x * y + x * z + y * z$ does not change the locations of the stationary points. That is to say, one can accomplish this circuit with three proteins, with two binding sites each, in which the occupation of either would enable their own transcription; also the z protein destroys x , y protein destroys z , and x destroys y .

To be able to fully fit the experiments of [St. Pierre \[2009\]](#) the phase space would partition into more than just two states, but four states, in addition to a splitting of a cell population into lysis and lysogeny, some cells (refractory) were immune to the phage and a small fraction of the lysogens had daughter cells that lyse. The number and properties of lysogens that had daughters that underwent lysis (5/100) could be useful in additional

studies of robustness¹. We continue the analysis of [Zhu et al. \[2004\]](#), [Wang et al. \[2010\]](#). Including the O_L , RNAP, tetramers, etc. could contribute this additional complexity, codes are provided [Borggren \[2011\]](#) if one was inclined towards such study.

At different times in the cell cycle different genes are turned on or off. Often these genes are transcripts for proteins that are necessary for functions in that particular interval. The concern here is with the lysogeny decision and in this interval we have only the O_R , cI gene, and Cro gene on.

Letting x , y denote the cI and Cro monomer numbers respectively we write for the dynamics

$$\dot{x} = \nu_x Z_x(x, y) / Z(x, y) - x / \tau_x, \quad (3.6)$$

$$\dot{y} = \nu_y Z_y(x, y) / Z(x, y) - y / \tau_y. \quad (3.7)$$

The creation term in the dynamics of the associated protein populations within a cell are governed by the rates at which RNA polymerase binds and translates the DNA into mRNA, together with the rate at which the ribosome translates the mRNA into the chain of amino acids that comprise the protein. In this way a cell seems to possess three different molecules encrypting the same message. Categorically speaking, the fundamental biological information processing is governed by the physics of $RNAP : DNA \rightarrow RNA$, $Ribosome : RNA^3 \rightarrow amino\ acid$. The net effective rate of these processes enters directly into the equations of motion through the $\nu_{x,y}$ variables.

One can say much about these questions without specifying numbers of binding sites or which promotional states enter into $Z_i(x, y)$ if we grant a few specifications. Firstly, we assert the positivity of τ_x and ν_x . This is also largely natural selections choice, though alternatives are imaginable. For example, RNAP can indeed unsubscribe a transcript, the backward reaction is permissible, and it is said that for any 11 forward reactions, there were 10 steps back [Bennett \[1982\]](#). The partition functions are positive by virtue of being well-behaved non-negative partition functions or a subset thereof. To avoid singularities and to insure $\dot{x}_i > 0$ we include at least one state that does not depend on x and y in each of Z , Z_x , and Z_y . In the case of λ this is provided by the open DNA configuration. This is the reference state $\Delta G_{000} = 0$. In the original formulation, [Shea and Ackers \[1985\]](#), security against singularities was also provided by the explicit inclusion of $[RNAP]$ and $[RNAP]^2$ states. For two-dimensional systems, for example two genes expressing two proteins, this is sufficient to prove the existence of any non-negative number of stationary points provided a sufficient supply of binding sites.

Stationary states occur when the derivatives vanish:

$$xZ(x, y) - \nu_x \tau_x Z_x(x, y) = 0, \quad (3.8)$$

$$yZ(x, y) - \nu_y \tau_y Z_y(x, y) = 0, \quad (3.9)$$

¹An interesting explanation without requiring more basins of attraction would be, if, for example, as a cell divides the daughter cells have each taken half the cI dimers from the parent lysogen, and in so doing switch to the lytic basin of attraction, causing the cells to lyse. (It is imaginable that whereas x , y may sit in one basin of attraction $x/2$, $y/2$ may reside in the other)

This is to illustrate some of the most important aspects of the equation and how these states interplay to create the desired states.

When x, y are zero, \dot{x}, \dot{y} are positive since $\nu_{x,y}$ and $Z_x(x, y)/Z(x, y)$ are. As x, y go to inf the $-x/\tau_x$ and $-y/\tau_y$ terms dominate and since $\tau_{x,y}$ are positive, the dynamics are confined to the positive quadrant. It follows similarly for y as well as x that there exists a contour for which $\dot{x} = 0$ on their respective axis since $0 < Z_x(x, y)/Z(x, y) \leq 1$; thus \dot{x} switches eventually to negative and stays negative and this crossing requires at most $\nu_x \tau_x$ proteins.

This provides an experimental test, or at least an avenue into parameterizing the models. For example, if one were to measure only $\nu_{x,y}$, $\tau_{x,y}$, and just the expectation value of $Z_{x,y}(x, y)/Z(x, y)$, one already has means to predict the average number of proteins x and y . Since the expectation value of $Z_{x,y}(x, y)/Z(x, y)$ is just the fraction of time RNAP is promoting, it is a much easier process than discovering all possible states, measuring their relative affinities and then simulating.

It is the fitting of empirical models to experimental data that comprises the bulk of the quantitative features of biophysics.

With cI at 236 amino acids, requiring at minimum 708 base pairs and Cro, 66 amino acids for 198 pairs, it is not surprising that Cro can be transcribed faster. The decay time in seconds is $\tau_{cI} = 2943$ and $\tau_{Cro} = 5194$.

In this model the states that contribute to the numerator for x are those in which O_{R3} is unoccupied and thus open for RNAP binding. Of these states those with a cI dimer bound at O_{R2} are believed to further enhance the rates. Thus,

$$\begin{aligned} \frac{cI}{dt} = & \frac{1}{Z} (\\ & mpi * Trm * Eci ([cI] e^{-\frac{\Delta G_{010}}{RT}} + [cI]^2 e^{-\frac{\Delta G_{110}}{RT}} + [cI][Cro] e^{-\frac{\Delta G_{210}}{RT}}) \\ & + Trmu * Eci (1 + [cI] e^{-\frac{\Delta G_{100}}{RT}} + [Cro] (e^{-\frac{\Delta G_{020}}{RT}} + e^{-\frac{\Delta G_{200}}{RT}}) + [cI][Cro] e^{-\frac{\Delta G_{120}}{RT}} + [Cro]^2 e^{-\frac{\Delta G_{220}}{RT}}) \\ &) - \frac{cI}{\tau_{cI}} \end{aligned} \quad (3.10)$$

$$\begin{aligned} \frac{Cro}{dt} = & \frac{1}{Z} (\\ & mpi * Trr * Ecro * (1 + [cI] e^{-\frac{\Delta G_{001}}{RT}} + [Cro] e^{-\frac{\Delta G_{002}}{RT}}) \\ &) - \frac{Cro}{\tau_{Cro}} \end{aligned} \quad (3.11)$$

Here T_{RM}, T_{RMU} , and E_{cI} are 0.115 cI monomers/sec, 0.01045 cI monomers/sec and 1 copy per mRNA synthesized. Next, RNAP is assumed to require both O_{R1} and O_{R2} free to bind to the Cro promotion site. This occurrence is repressed in comparison with the states that promote cI but promotions proceed with a higher burst. T_{RR}, E_{Cro} are 0.30 Cro monomers/sec and 20 copies per mRNA synthesized, respectively. mpi denotes the number of phage O_R sites per infected cell. The mpi factor curiously seems to only have

been multiplied by the premotional states with cI dimer at O_{R2} and has been given the value 3. Here, Z is the partition function from 2.1, while $[cI]$ and $[Cro]$ are the assumed relations in 2.3 and 2.4 respectively. Again, these are the assumptions used to correspond to Zhu et al. [2004], Wang et al. [2010] and to references therein. We will posit different equations of motion, including the correction to the *pnc* confusion, and the effect of RNA polymerase in *variations on a theme*. Experimental collaboration, references, or data to analyze that can pin down any to all parameters are welcome, appreciated, and daresay necessary.²

The situation is more complicated in that cI promotion occurs at 2 different rates, the states with a cI dimer on O_{R2} enhance the actual rate which RNAP transcribes, as well as binding more favorably, Ptashne [2011], and how this assistance translates physically into the enhanced rate is an interesting open question. Here it is expressed by using the two different transcription rates, T_{RM}, T_{RMU} .

To approximate the effects of specific mutations we consider mutations, λ_{ijk} , by which it is meant that new cells have been synthesized with $O_{R1} \rightarrow O_{Ri}$, $O_{R2} \rightarrow O_{Rj}$ and $O_{R3} \rightarrow O_{Rk}$, the chemical mutations correspond to changes in the strings shown in Fig. 1.4 which lead to changes in the binding affinities in Table 2.1.

3.2 Stationary Points

The stationary points of the system are found empirically and tabulated herein.

To inquire into the nature of these stabilities it is fruitful to consider the linearized equations of motion at a stable point which approximates the flow in the neighborhood. Stable, unstable, and saddle points have negative, positive, and mixed eigenvalues for their respective linearized equations. An imaginary part suggests the dynamics around the point have non vanishing circulation.

Although we will maintain a global perspective in the Diffusion section to follow, in correspondence with the main utility of the methods in Wang et al. [2008] or Wang [2011] where a local analysis is not necessary before proceeding to find global features, dynamics near stationary points is mathematically very well understood Perko [2001] and even with stochastic features Kwon et al. [2005], and any such analysis would begin with an enumeration of the fixed points provided here. It is also noticed that the peaks of the distributions to be found are displaced from the fixed points of the underlying dynamical system. This displacement is purely the result of the inclusion of noise and has been observed in theory and numerical studies before Kwon and Ao [2011]. An experimental situation where the classical dynamics can be fixed and the noise part can be scanned independently of these dynamics could greatly facilitate the observations and interpretations of the non-equilibrium steady states and the breaking of detailed balance suggested in Wang et al. [2008], Wang [2011], Qian [1998], Ao [2004] by directly measuring this displacement.³

²nborggren@gmail.com

³Doing this in a minimally biased fashion could be difficult. For example, scanning noise by simply letting the number of particles decrease can easily put one in a domain where the classical kinetics are no longer a good approximation Samoilov and Arkin [2006]. In such an instance the measured displacement, in addition to the contribution from noise, would also include a

For these systems the relevant matrix components are, writing x_1, \dots, x_n as x ,

$$\begin{aligned} \mathcal{D}f(x_1, \dots, x_n)_{ij} = & \frac{\nu_{x_i}}{Z(x_1, \dots, x_n)^2} [Z(x_1, \dots, x_n) \frac{\partial}{\partial x_i} Z_{x_j}(x_1, \dots, x_n) \\ & - Z_{x_j}(x_1, \dots, x_n) \frac{\partial}{\partial x_i} Z(x_1, \dots, x_n)] \\ & - \frac{\delta_{ij}}{\tau_{x_i}}, \end{aligned} \quad (3.12)$$

where δ_{ij} is one when $i = j$ and is otherwise zero while the partition functions are evaluated at the stationary point under query.

For the two by two case the eigensystem is

$$\alpha_1 = \frac{1}{2}(-\sqrt{a^2 - 2ad + 4bc + d^2} + a + d), \quad (3.13)$$

$$e_1 = (-\frac{\sqrt{(a^2 - 2ad + 4bc + d^2)} - a + d}{2c}, 1), \quad (3.14)$$

$$\alpha_2 = \frac{1}{2}(\sqrt{a^2 - 2ad + 4bc + d^2} + a + d, \quad (3.15)$$

$$e_2 = (\frac{\sqrt{(a^2 - 2ad + 4bc + d^2)} + a - d}{2c}, 1), \quad (3.16)$$

with

$$a = \frac{\nu_x}{Z^2} [Z \frac{\partial}{\partial x} Z_x - Z_x \frac{\partial}{\partial x} Z] - \frac{1}{\tau_x}, \quad (3.17)$$

$$b = \frac{\nu_x}{Z^2} [Z \frac{\partial}{\partial y} Z_x - Z_x \frac{\partial}{\partial y} Z], \quad (3.18)$$

$$c = \frac{\nu_y}{Z^2} [Z \frac{\partial}{\partial x} Z_y - Z_y \frac{\partial}{\partial x} Z], \quad (3.19)$$

$$d = \frac{\nu_y}{Z^2} [Z \frac{\partial}{\partial y} Z_y - Z_y \frac{\partial}{\partial y} Z] - \frac{1}{\tau_y}. \quad (3.20)$$

In two dimensions the quantity $(ad - bc)$ is the determinant. The fact that this term need not vanish is why the notion of energy landscape must be generalized into the flux framework to include fields that are not the gradient of a potential, see [Wang et al. \[2008\]](#).

Understanding these equations can tell us a great deal about the type of forces at work since near a particular point, $x' = (x_1', \dots, x_n')$ through an expansion [Perko \[2001\]](#), we know that the system is qualitatively similar to the linearized equations

$$\dot{x} \approx f(x') + \mathcal{D}f(x')x. \quad (3.21)$$

contribution from reality deviating from the kinetic assumptions of the model.

If $\alpha_1 < 0$, $\alpha_2 < 0$ then the point serves as a probability sink, if $\alpha_1 > 0$, $\alpha_2 > 0$, then it is a probability source. Mixed eigenvalues are transition states. For example, one can verify that any of the following conditions lead to stability:

$$a < 0, b = 0, d < 0$$

$$a = 0, b > 0, c < 0, d \leq -2\sqrt{-bc}$$

$$a = 0, b < 0, c > 0, d \leq -2\sqrt{-bc}$$

$$a > 0, b > 0, c < -\frac{a^2}{b}, \frac{bc}{a} < d \leq a - 2\sqrt{-bc}$$

$$a > 0, b < 0, c > -\frac{a^2}{b}, \frac{bc}{a} < d \leq a - 2\sqrt{-bc}$$

$$a < 0, b > 0, c \geq 0, d < \frac{bc}{a}$$

$$a < 0, b > 0, -\frac{a^2}{b} < c < 0, a + 2\sqrt{-bc} \leq d < \frac{bc}{a}$$

$$a < 0, b > 0, -\frac{a^2}{b} < c < 0, d \leq a - 2\sqrt{-bc}.$$

Of the 27 mutations considered, 15 had underlying dynamical systems with 3 stationary points, 11 had one stationary point, and one had two stationary points. The locations are tabulated in the tables A.1, A.2, A.3 and A.4. In particular we note that the permutation group S_3 acting on O_{R1}, O_{R2} , and O_{R3} , only six mutations, is sufficient to provide mutants with a single lytic peak λ_{231} , λ_{321} a single lysogenic peak λ_{132} while $\lambda_{123}, \lambda_{312}$, and λ_{213} retain the switch properties.

The 27 phase portraits of this model can be seen in Tables A.1, A.2, A.3. Dynamics are also combined with the results of the diffusion section to demonstrate the importance of statistical, trajectory based approaches. Those results are in the tables A.25, A.26, A.27 and demonstrate that the different limit cycles correspond to regions of different entropy or energy density.

This model, however, does not probe what the downstream reactions do to the phase space distribution given to it; it is quite likely that in the case of, for example, λ_{231} lysogens will still attempt to form but with a loss of stability. It does, however probe transient phenomena and the initial impulse of the system is, even in the monostable case, towards two states.

$$\begin{aligned} \frac{\partial Z_x}{\partial x} &= \frac{\partial [cI]}{\partial x} (e^{-\Delta G_{010}} + e^{-\Delta G_{100}}) + \\ &2 * [cI] * \frac{\partial [cI]}{\partial x} e^{-\Delta G_{110}} + \frac{\partial [cI]}{\partial x} [Cro] (e^{-\Delta G_{210}} + e^{-\Delta G_{120}}) \\ \frac{\partial Z_x}{\partial y} &= \frac{\partial [Cro]}{\partial y} (e^{-\Delta G_{020}} + e^{-\Delta G_{200}}) + \\ &[cI]^2 e^{-\Delta G_{110}} + \\ &2[Cro] \frac{\partial [Cro]}{\partial y} e^{-\Delta G_{220}} + \\ &[cI] \frac{\partial [Cro]}{\partial x} (e^{-\Delta G_{210}} + e^{-\Delta G_{120}}) \end{aligned} \quad (3.22)$$

$$\frac{\partial Z_y}{\partial x} = \frac{\partial [cI]}{\partial x} e^{-\Delta G_{001}}, \quad \frac{\partial Z_y}{\partial y} = \frac{\partial [Cro]}{\partial y} e^{-\Delta G_{002}} \quad (3.23)$$

The robustness of the lysogenic cycle would be unsurprising if the system were wholly deterministic. Dynamical systems in general can partition into multiple basins of attraction and thus so too can those arising from chemical kinetics. In two dimensions, like that presumed here, it can often be the case that there exists some contour that partitions the phase space [Perko \[2001\]](#). This contour is unstable in the perpendicular direction and deterministic trajectories beginning on this contour take the steepest descent to a saddle point if the system is bounded. Any crossing from one region to the other would require an intersection with this line and is forbidden with the rules for ordinary differential equations. Thus, hopping would require a stochastic process to change this fact for this model. The puzzle of this stability, which we consider in the next section, is why there is not enough noise.

4 Diffusion; Evolution of Ensembles

Lo otro no existe: tal es la fe racional, la incurable creencia de la razón humana. Identidad=realidad, como si, a fin de cuentas, todo hubiera de ser, absoluta y necesariamente, uno y lo mismo. Pero lo otro no se deja eliminar; subsiste, persiste; es el hueso duro de roer en que la razón se deja los dientes.

Antonio Machado, *Juan de Mairena*, quoted in *El laberinto de la soledad*, Octavio Paz

When Galileo first assisted his gaze with the telescope and directed it towards Saturn, he discovered its ring, and Titan, one of its moons. The point-like orb of the ancients revealed more structure still when Cassini subsequently fixed his gaze upon this ring, revealing a large division; the ring was not one, but two. Further query revealed yet more structure, courtesy of an *in situ* spacecraft, also named Cassini, and one is now disinclined to the naming of rings, which appear with the abundance of grooves in a phonograph.

Indeed the rings of Saturn are not rings at all, but a collection of granules of various sizes caught in circular potential wells; the only rings are in the energy landscapes, our abstractions, and only appear in the data in that there was a sufficient supply of granules to sample the space so that the typical distance between stones was shorter than the resolution of our instruments.

With that said let us revisit the diffusion equation

$$\partial_t u = \partial_i (D_{ij} \partial_j u - F_i u). \quad (4.1)$$

For, example if one were to construct a Hamiltonian of the Saturn system and fill a vector, F_i , with each of the equations of motion, then u would denote the density of granules, and in the case of D_{ij} being zero, the equation would describe the evolution of that initial configuration as it samples the Saturn force field; depending on the details some fraction fall into Saturn's core, some might fall into Titan's surface, others escape, and some get trapped into the resonances of the system, the rings. Letting the $D > 0$ would allow our granules to interact with each other, although they are small enough so as not to perturb F_i , they are of comparable sizes to each other and collisions lend them to diffuse through space.

In the case of λ phage, the numbers of proteins involved are small, and intrinsic noise, shot noise, is indeed a real source of diffusion for the system; the system kinetics do not behave the same with every trial, but spread throughout the phase space. However, Eq. (4.1), can equally describe such a situation. For references and the general state of affairs of Fokker-Planck equations used as a *theory of everything*, query [Friston and Ao \[2012\]](#).

4.1 Master Equations, Langevin Dynamics, or Diffusion Equation

Many approaches are used to capture the chemical kinetics or gene expression in a noisy environment, all of which are phenomenological and have their advantages and disadvantages. The proper approach can vary from system to system and depend on resources available. See for example [Wang et al. \[2008\]](#), which this thesis follows and extends with the finite elements, and also the potential approaches in [Ao \[2004\]](#), [Zhu et al. \[2004\]](#); for field-theoretic approaches there is [Wang and Wolynes \[1996\]](#), [Wang et al. \[2010\]](#), [Feng and Wang \[2011\]](#), and for Langevin dynamics and master equations see [Wang et al. \[2006\]](#), [Ge and Qian \[2010\]](#), [Qian et al. \[2002\]](#).

4.1.1 An Integral Approach for Diffusion Equations and Passage Times

The finite element method is emerging as a universal method for the solution of differential equations of arbitrary complexity from all walks of science, see [Logg et al. \[2012\]](#). The following derivation is typical of the procedures implemented throughout the FEniCS book ([Logg et al. \[2012\]](#)). In order to use a finite element method to solve partial differential equations it is necessary to first translate the equations into an integral equation. This integral equation can be seen as setting a bilinear functional equal to a linear functional and solving the corresponding system by standard linear algebraic methods. It will be described in detail for the Fokker-Planck equation and corresponding passage time equations.

4.1.1.1 Diffusion Equations

First we seek to solve:

$$\partial_t u = \partial_i (D_{ij} \partial_j u - F_i u). \quad (4.2)$$

For numerics the function space is specified by defining a mesh over Ω and the type of element associated with each point in the space. These specifications alter the analysis mainly in how long it takes for a desired level of accuracy to be reached. We continue thinking of the function space as the space of twice differentiable functions over Ω . The functional form is found by multiplying a test function v and integrating over the mesh

$$\int_{\Omega} v \partial_t u dx = \int_{\Omega} v \partial_i (D_{ij} \partial_j u - F_i u) dx. \quad (4.3)$$

Continuing with integration by parts yields

$$\int_{\partial\Omega} v (D_{ij} \partial_j u - F_i u) n_i ds - \int_{\Omega} \partial_i v (D_{ij} \partial_j u - F_i u) dx. \quad (4.4)$$

Here D_{ij} denotes the spatial dependance of the Diffusion, F_i , is the drift force, n_i , a normal vector to the mesh. The mesh is denoted Ω with boundary $\partial\Omega$.

We approximate to find an expression for $u_{t+\delta t}$ given the solution at u_t

$$\begin{aligned} \int_{\Omega} v(u_{t+\delta t} - u_t) dx \approx & \delta t \left[\int_{\partial\Omega} v(D_{ij}\partial_j u - F_i u) n_i ds \right. \\ & \left. - \int_{\Omega} \partial_i v(D_{ij}\partial_j u - F_i u) dx \right]. \end{aligned} \quad (4.5)$$

We take the average of the expression over time δt by asserting $u \approx \frac{u_{t+\delta t} + u_t}{2}$. Continuing gives

$$\begin{aligned} & \frac{\delta t}{2} \left[\int_{\partial\Omega} v(D_{ij}\partial_j u_{t+\delta t} - F_i u_{t+\delta t}) n_i ds \right. \\ & - \int_{\Omega} \partial_i v(D_{ij}\partial_j u_{t+\delta t} - F_i u_{t+\delta t}) dx \\ & + \int_{\partial\Omega} v(D_{ij}\partial_j u_t - F_i u_t) n_i ds \\ & \left. - \int_{\Omega} \partial_i v(D_{ij}\partial_j u_t - F_i u_t) dx \right]. \end{aligned} \quad (4.6)$$

Collecting terms with $u_{t+\delta t}$ on one side of equality and u_t on the other gives the variational forms a and L that we will use

$$a = \int_{\Omega} (v u_{t+\delta t} + \frac{\delta t}{2} \partial_i v [D_{ij}\partial_j u_{t+\delta t} - F_i u_{t+\delta t}]) dx, \quad (4.7)$$

$$L = \int_{\Omega} (v u_t - \frac{\delta t}{2} \partial_i v [D_{ij}\partial_j u_t - F_i u_t]) dx. \quad (4.8)$$

$a(v, u) = L(v)$ then defines a linear system. To see this note that if v, v' satisfy a then so does $\alpha(v + v')$ and similarly for L for all scalars α . Since it also holds that a is linear in $u_{t+\delta t}$, a is bilinear. The u_t in L is specified by propagating u_0 and thus explicitly appears as a known function on the right side of the equations. Thus, finding the u that solves $a(v, u) = L(v)$ for all v advances our solution a timestep. Eq. (5.1) shows the simplicity of the lines of code to perform these calculations. Indeed the computation is more general and concise then the notation here.

The passage time equations are much simpler. The integral equation corresponding to $F_i \partial_i t_1(x) - D \partial_i t(x) \partial_i t_1(x) = -1$ follows by simply multiplying through by the test function and integrating

$$\int_{\Omega} (t(x)(F_i \partial_i t_1(x)) - D \partial_i t(x) \partial_i t_1(x)) dx = - \int_{\Omega} t(x) dx, \quad (4.9)$$

$$\int_{\Omega} (t(x)(F_i \partial_i t_n(x)) - D \partial_i t(x) \partial_i t_n(x)) dx = -n \int_{\Omega} t(x) T_{n-1} dx. \quad (4.10)$$

One can then find the passage times recursively. The induction step is bilinear provided one has first solved for $T_{n-1}(x)$ before proceeding to find $T_n(x)$.

To perform the simulations in this work we need still to specify a mesh and type of finite element. We choose for Ω the domain in protein number space $0 < cI < 1000$ and $0 < Cro < 1000$. We discretize this into at least a 50 by 50 grid and use a quadratic

Lagrange element. The quadratic Lagrange elements span the surface by connecting the adjacent mesh points with a quadratic function (see Fig. 4.1). The function space is the set of all such functions. One benefit of a finite element method over a finite difference is the functions in the function space are continuous over Ω and we can evaluate any function, and in particular our solutions, anywhere in the domain. We see this smoothness manifest in the plots which evaluate the functions along the trajectories.

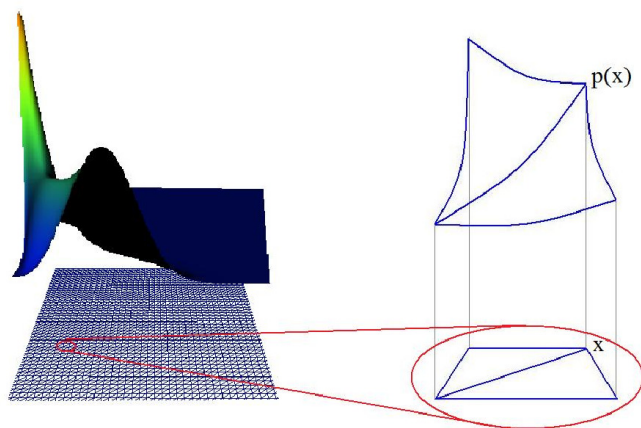


Figure 4.1: **Finite elements construct a surface by combining simplices together. The quadratic elements used fit the function inbetween the mesh points with a quadratic formula.**

4.1.2 On Convergence

Technically, the asymptotics of the Fokker-Planck equation take infinitely long to converge. Such a program, see Fig. 4.2, can still be written, albeit not technically admissible into von Neumann architectures. Taking the a and L as bilinear and linear forms, respectively, defined anywhere, we have an infinite trajectory pursuing eternally its asymptotic states. A break can be entered by establishing a criterion for convergence. The distributions present in the Appendix were found using 53, 7.5 second time steps and well approximate the steady state solutions.

This loop was carried out 53 times before I choose to break it in the states labeled steady in this work, with $k = 7.5$ 'seconds' the time step.

4.1.3 Choice of Gauge

It is noticed that equation (4.1) presupposes a choice of gauge and the curl of any vector field can be added to the flux landscape, $D_{ij}\partial_j P - F_i P + \epsilon_{ijk}\partial_j A_k$, and still satisfy the basic equation. We choose a gauge such that $\epsilon_{ijk}\partial_j A_k = 0$. We notice that what has been called the entropy production in Wang et al. [2008] is a gauge-dependent quantity and how it translates into entropy or heat in absolute units is open. We note that the different kinetic paths will have different thermodynamic signatures, and understanding the connection can enable direct measurement of some of the quantities that have been proposed.

A recent paper has elaborated on the nature of the relationship between related Fokker-Planck equations and gauge theories, Feng and Wang [2011], this is likely to prove a rather pregnant approach to chemical kinetics.

Figure 4.2: This would be an example of a legitimate python code that runs indefinitely. Computationally, the asymptotics of the equations require a function that does not terminate.

Python code

```
def TimeStep(u0,u1):

    A = assemble(a)

    while 'no your solutions have not converged':
        b = assemble(L)
        bc.apply(A,b)
        solve(A, u1.vector(), b)
        u0.interpolate(u1)
        yield u1

while InfiniteTraj:InfiniteTraj.Next()
```

Figure 4.3: Calculating expectation values is straightforward, even on a command line.

Python code

```
>>> exprs = [Expression("x[0]"),Expression("x[1]"),
... Expression("x[0]*x[0]"),Expression("x[1]*x[1]")]
>>> ans=[
...
    assemble(i*u*dx,mesh=mesh)/assemble(u*dx,mesh=mesh)
    for i in exprs]
>>> ans
[93.076476131804114, 169.77672009751805,
 13624.244707344435, 36974.455745161322]
>>> sqrt(abs(pow(ans[0],2)-ans[2]))
70.434468112069482
>>> sqrt(abs(pow(ans[1],2)-ans[3]))
90.279128585129428
```

4.1.4 $\Delta x \Delta J$

After having computed the solutions to the equations we can calculate some moments of the distribution. So far the diffusion has been picked to correspond to the numbers believed to quantify robustness, see [Zhu et al. \[2004\]](#), [Wang et al. \[2010\]](#), however the boundary condition can be tuned to make this number vary over many orders of magnitude and it may be more prudent to use moments of the distribution to parameterize D . For example when cI proteins are reported in numbers 150-250 in a cell this can suggest an expectation value of, say, 200 for x and a spread of 50 for the second moment. These numbers are more easily measureable then the robustness, which has been the subject of amendment over time.

The figure 4.3 illustrates how FEniCS software can compute expectation values from the command line, and the type of calculations that can be performed once one has solutions in hand. Let us compute Δx and Δy with the distributions from A.10 and recompute the related expectation values.

It is important to realize that it is not the mass of an electron that biologists measure. The errors do not necessarily go away with better measurements. The numbers are

supposed to fluctuate and this can actually be interpreted to mean that the higher moments of the distribution are relevant. A spread of 70 in the number of *cI* proteins and 90 *Cro* proteins is predicted by this model.

4.2 Passage Times

The equations again for the first passage time is

$$\int_{\Omega} (t(x)(F_i \partial_i t_1(x)) - D \partial_i t(x) \partial_i t_1(x)) dx = - \int_{\Omega} t(x) dx, \quad (4.11)$$

where we force a boundary condition of zero, in the present case a circle of radius ten is placed around the critical point, where we want to calculate the passage time to and following we tabulate the results for the mutants we can construct from the binding values available. First, we explore the properties of the 15 of 27 mutants that maintain the switch properties in this model. We notice however that the illusion of bistability is manifest in the transient phenomena even for the monostable states. That is to say, if the switch behavior is a non-equilibrium phenomenon, monostable mutations may be relevant.

4.3 Noise-induced stability

The dynamics picture of the previous chapter is quite idealized. Even if all the parameters were perfectly known there is still the issue that we are discussing mesoscopic regimes and our large N intuitions and classical assumptions are certain to break down in this limit. The constant rate assumptions would need to be replaced with discrete events and the manner in which a unit of time splits into subunits becomes important, as well as the annihilation processes. The changes to the partition function in this model are again relevant.

It is found here that the noise added displaces to a large extent the locations of the peak from the corresponding stable points of the underlying dynamical system. Viewing the noise as diffusion through concentration space and developing effective field theories and path integrals has provided interesting and useful constructions, see [Wang et al. \[2010\]](#). The fact that there even exist corrections to be made is interesting. This is the type of functionality that Schrodinger deliberates on being prohibited for an organism, in his discussion of the $\sqrt{(n)}$ law.

... an organism must have a comparatively gross structure in order to enjoy the benefit of fairly accurate laws, both for its internal life and for its interplay with the external world. For otherwise the number of co-operating particles would be too small, the 'law' too inaccurate. The particularly exigent demand is the square root. For though a million is a reasonably large number, an accuracy of just 1 in 1,000 is not overwhelmingly good, if a thing claims the dignity of being a 'Law of Nature.'

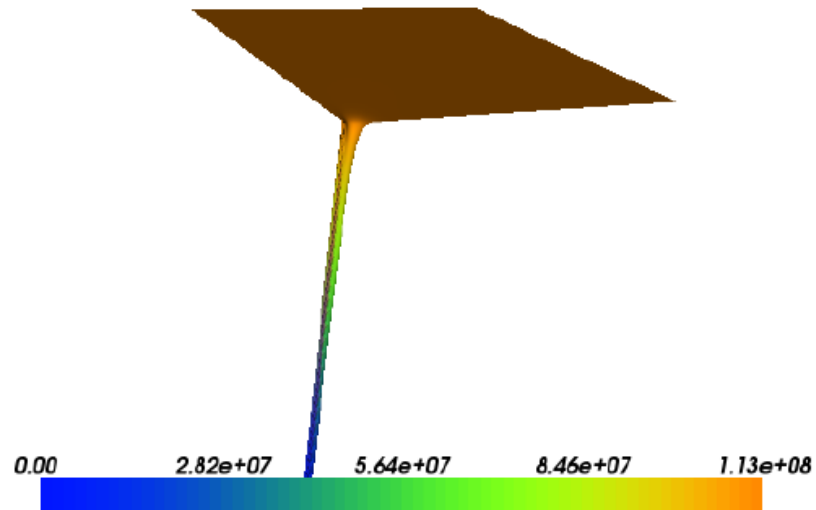
Erwin Schrödinger 'What is Life'

For λ phage this is just a restatement of the stability puzzle: how then can the numbers of proteins be within the range 0-1000, a far worse scenario than Schrodinger deemed

Figure 4.4: The first passage time with an absorbing boundary condition within 10 units of the origin. The origin is extremely unstable; The steepness of the funnel of the first passage time landscape illustrates the overwhelming propensity for the system to be driven towards some non-trivial state. The numbers, $\approx 10^8$, appear in seconds.

Python code

```
...
Assembling matrix over cells [=====
=====] 100.0%
Applying boundary conditions to linear system.
Solving linear system of size 1442401 x 1442401
(PETSc LU solver, umfpack).
*** Warning: Using LU solver, ignoring preconditioner
"default".
evaluate at wells, [0.14175299999999999,
462.57299999999998] 112474076.373
evaluate at wells, [644.052000000000002,
0.0061165799999999999] 112473547.501
evaluate at wells, [29.9798000000000001,
178.22999999999999] 112469292.141
```



possible? *Why* does a deterministic dynamical analysis in the small concentration regime *work at all*? Do we retain the observed robustness when we consider noise? Is the noise *really* $\sqrt{(n)}$ noise? What paths are permissible and what are their relative probabilities? General questions and methods of this variety are the subject of intense debate and muse to much creativity, see Wang and Wolynes [1996], Wang et al. [2010], Stock et al. [2008], Ghosh et al. [2006], Qian et al. [2002], and have been applied to the λ system, see Zhu et al. [2004], Wang et al. [2010]. This debate has led to legitimate concerns and elegant measurements to seek explanations from other variables, like volume, where the variation across an ensemble are less severe or can be experimentally minimized, see St. Pierre [2009].

It is written that $D_{cI,Cro} = \text{const} * \tau_{cI,Cro} / N_{cI,Cro}$ then $\Lambda = D^{-1}$, Zhu et al. [2004], Wang et al. [2010]. However, their expression for Λ is the same as their expression for D . The

calculations require

$$D = \begin{pmatrix} 17.85 * \frac{N_{cI}}{\tau_{cI}} & 0 \\ 0 & 25.0 * \frac{N_{Cro}}{\tau_{Cro}} \end{pmatrix}. \quad (4.12)$$

An $N_{cI} = 644.052$ corresponds to the value at the lysogen zero and $N_{Cro} = 462.573$ the lytic zero. $\tau_{cI} = 2943s$ and $\tau_{Cro} = 5194s$ giving

$$D_{wt} \approx \begin{pmatrix} 3.91 & 0 \\ 0 & 2.2271 \end{pmatrix}. \quad (4.13)$$

The numerics proceed with this diffusion tensor. If the analogy is to extend from diffusion in ordinary space to diffusion in concentration space, then the mean squared displacement is a displacement in concentration space and the units of this D should be $dN_{cI}dN_{Cro}/ds$.

Let us attempt to motivate the form of Eq. (4.12) and the scale of Eq. (4.17). If

$$x \Rightarrow x \pm \sqrt{x} \quad \text{then} \quad \dot{x} \Rightarrow \dot{x} \pm \frac{\dot{x}}{2\sqrt{x}} \approx f(x) \pm \frac{f(x)}{2\sqrt{x}}. \quad (4.14)$$

For large x , we have seen that the term $-x/\tau$ dominates and recover fluctuations of order \sqrt{x} . Since D goes as the square of these fluctuations we get the proportionality to x as appears in 4.12. The situation is more subtle when x is small. For Shea and Acker type models the D_{xx} term, say, would be dominated by terms of the form Z_x^2/xZ^2 which can diverge. We expect corrections from this end to increase the average diffusion and therefore facilitate the crossing of a barrier, thereby reducing the robustness of the switch more than Eq. (4.17) may suggest.

The cI protein is found in numbers typically less than 180-350 for a healthy cell, and error from counting statistics alone contributes $\pm\sqrt{(N)}$ to the intrinsic noise. The uncertainty in ΔG and the dimerization equation of state also induces a spread of trajectories. A convenient way to catch these and other sources of stochasticity is to assert a diffusion tensor, D_{ij} , with i, j twice indexing over the protein species. This is equivalent to considering some Langevin-like equation but allows us to use partial differential equations to explore ensembles of trajectories. Noise of greater complexity than Gaussian noise can be readily studied with the software developed.

As with the other parameters in this model they are left determined by experiments and have been given a simple form to understand the effects that chance events could have on the dynamical picture. In particular, it can wash away higher resolution features which can, for example, remove the gap between two nearby stabilities, facilitating perhaps the appearance of bistability when the dynamical picture may suggest more structure. From the modeling perspective it can be useful to use large, perhaps unphysically large (analogous to a high temperature limit), diffusion to speed up simulations and see some of the more basic underlying features of the phase space. As diffusion gets smaller, the steady states take exponentially longer to converge and can quickly become computationally prohibitive. In these scenarios, where converging to a steady state from any arbitrary initial condition is computationally prohibitive, the steady states from larger diffusion cases can serve as useful initial conditions since they have already undergone a large degree of this convergence process and the remaining time to equilibrium is for the finer resolution

features to develop since the coarse features were already present at the beginning of the simulation.

4.4 Little's Mutants

Mutations in the O_R region have been performed, see [Little et al. \[1999\]](#), and we would like to check the performance of our model and suggest means to calibrate accordingly. The qualitative features compare to experiment but quantitative comparison is premature; in particular, the time and concentration units need to be understood better. As is, the first passage times predict a smaller barrier then observed.

Diffusion Tensor		
mutant	D_{xx}	D_{yy}
$\lambda_{123}wild\text{-}type$	3.90778	2.22668
λ_{121}	0.775855	2.22668
λ_{323}	2.6462	1.5024
$\lambda_{323'}$	2.0016	2.86862

Table 4.1: Values for diffusion tensors. λ_{323p} in [Little et al. \[1999\]](#) is nearly λ_{222} , the cI affinities are identical and Cro values differ by -6 kcal/mol in the first and third site.

4.4.1 Wild-Type Evolution

Four trajectories of distributions are depicted in Tables [A.5](#), [A.6](#), [A.7](#) and [A.8](#), corresponding to the wild-type and mutants λ_{121} , λ_{323} , and $\lambda_{323'}$. It is evident that in the first 200 seconds, an initially peaked distribution has developed a second peak and after about 600 seconds the distributions have reached their steady states.

4.4.2 Steady state properties for O_R Mutations

The Tables [A.9](#), [A.10](#), [A.11](#), [A.12](#), [A.13](#), [A.14](#), [A.15](#), [A.16](#), [A.17](#), [A.18](#), [A.19](#), [A.20](#), and [A.21](#) archive the results. Depicted are the phase space portraits alongside the Probability and Flux Landscapes and above the passage time distributions to each of the three stationary points. Assorted expectation values are calculated and can be compared with the wild-type.

In particular, we notice that with diffusion the peaks of the solutions are considerably displaced towards smaller numbers from where one would expect the peaks from the deterministic equations. This is required to satisfy the boundary condition, and it would be interesting to see how a Langevin-based approach can reproduce this ensemble property in that different trajectories would require correlation.

Noise in this model seems to destabilize at least the mutants λ_{111} , λ_{131} , and λ_{312} ; either no gap, or a small gap is present. The phage seems most robust in response to mutations in O_{R1} in that λ_{*23} all have a sizeable gap.

4.5 Hybrid particle and ensemble approaches

It is important to see how different paths contribute differently to the overall energy and entropy profiles, see [Seifert \[2005\]](#), [Stock et al. \[2008\]](#). To illustrate the shape of these landscapes it is sufficient to evaluate the steady state ensembles values along deterministic trajectories. These calculations are depicted in Tables [A.25](#), [A.26](#), and [A.27](#).¹ It is evident that different paths can converge to different locations and demonstrates the motivation for a (more fundamental) dynamic partition function.

4.6 Variations on a theme

We revisit the equations of motion with RNAP still included in the partition function. Here, the promotional states are the natural ones, the binding of RNAP to *prm* for cI promotion and the binding of RNAP to *pr* for Cro promotion. These promotional states enter the dynamics as follows.

$$\begin{aligned} \frac{d(cI)}{dt} = \frac{1}{Z} (& \\ & mpi * Trm * Eci([RNAP][cI]e^{-\frac{\Delta G_{013}}{RT}} + [RNAP][cI]^2e^{-\frac{\Delta G_{113}}{RT}} + [RNAP][cI][Cro]e^{-\frac{\Delta G_{213}}{RT}}) \\ & + Trmu * Eci([RNAP]^2e^{-\frac{\Delta G_{33}}{RT}} + [RNAP][cI]e^{-\frac{\Delta G_{103}}{RT}} + [RNAP][Cro](e^{-\frac{\Delta G_{203}}{RT}} + e^{-\frac{\Delta G_{023}}{RT}}) \\ & + [RNAP][cI][Cro]e^{-\frac{\Delta G_{123}}{RT}} + [RNAP][Cro]^2e^{-\frac{\Delta G_{223}}{RT}} + [RNAP]e^{-\frac{\Delta G_{03}}{RT}}) \\ &) - \frac{cI}{\tau_{cI}} \end{aligned} \quad (4.15)$$

$$\begin{aligned} \frac{d(Cro)}{dt} = \frac{1}{Z} (& \\ & mpi * Trr * Ecro * ([RNAP]e^{-\frac{\Delta G_{30}}{RT}} + [RNAP][cI]e^{-\frac{\Delta G_{31}}{RT}} + [RNAP][Cro]e^{-\frac{\Delta G_{32}}{RT}} + [RNAP]^2e^{-\frac{\Delta G_{33}}{RT}}) \\ &) - \frac{Cro}{\tau_{Cro}} \end{aligned} \quad (4.16)$$

The Z in these equations is now from [2.7](#), with values listed in table [4.2](#). The inclusion of these changes leads to much higher numbers for cI, up to 1500 proteins, though the lysogen peak still dominates. The time scale is slowed down in comparison to the dynamics without RNAP. The $[RNAP] = 600nM$ was chosen to be in the range of the experiment

¹[Seifert \[2005\]](#) suggests a definition of path entropy as $s(\tau) = -\ln((x(\tau), \tau))$, with $x(\tau)$ some stochastic trajectory. The definition is sufficient to define a path entropy for any function $x(\tau)$, regardless of the motivation of such functions. The principle of maximum caliber, see e.g. [Stock et al. \[2008\]](#), is an effort to define a dynamic partition function, over all such paths. There has been some success with this method for low number of states, but in the discrete case, or the continuous case here, the implications are largely unknown.

[Klumpp and Hwa \[2008\]](#) and is presumed to be a fixed number for the simulations. The results are displayed in tables [A.22](#) and [A.23](#).

An important contribution of the codes written in this thesis is the capacity for noise with spatial and temporal noise. As an example, a spatially dependant noise that distorts locally the classical kinetics is given in [4.17](#). A multiplier of 1000 is used to control the numerics.²

$$D(cI, Cro) = \begin{pmatrix} 1000 * cI^2 * \frac{(\frac{dcI}{dT})^2}{1+cI^2+Cro^2} & 0 \\ 0 & 1000 * Cro^2 * \frac{(\frac{dCro}{dT})^2}{1+cI^2+Cro^2} \end{pmatrix}. \quad (4.17)$$

izf

²Computationally, diffusion equations with finite elements requires a very fine grid, or a rotating coordinate system when the noise is smaller or comparable to the underlying flow. The factor of a thousand is an artificial enhancement to control the numerics. (i.e. a high temperature limit of sorts) In general, it is best to fit all parameters simultaneously with experiments.

λ_{123} , wild – type O_R affinities (kcal/mol)					ΔG_{ijk}		
protein	O_{R1}	O_{R2}	O_{R3}	ref.	equation	monomial	premotors
$\Delta G_{001}, \Delta G_{010}, \Delta G_{100}$	-15	-13	-12	Zhu et al. [2004]	$\Delta G_{000} = 0.0$	1	
$\Delta G_{002}, \Delta G_{020}, \Delta G_{200}$	-18.4	-17.1	-19.5	Zhu et al. [2004]	$\Delta G_{001} = -12.0$	[Ci]	
$\Delta G_{coop} = -6.9$					$\Delta G_{010} = -13.0$	[Ci]	
$pnc = 2.5 * 10^{-9}$					$\Delta G_{100} = -15.0$	[Ci]	
$[RNAp] = 300, 600nM$ Klumpp and Hwa [2008]					$\Delta G_{002} = -19.5$	[Cro]	
$\Delta G_{30} = -11.5, -10.5$, RNAP binding to Cro promoter.					$\Delta G_{020} = -17.1$	[Cro]	
$\Delta G_{03} = -11.5, -13.5$, RNAP binding to cl promoter.					$\Delta G_{200} = -18.4$	[Cro]	
$\Delta G_{ijk}(cont)$					$\Delta G_{011} = \Delta G_{010} + \Delta G_{001} + \Delta G_{coop}$	[Ci] ²	
					$\Delta G_{110} = \Delta G_{010} + \Delta G_{100} + \Delta G_{coop}$	[Ci] ²	
					$\Delta G_{101} = \Delta G_{100} + \Delta G_{001}$	[Ci] ²	
					$\Delta G_{022} = \Delta G_{020} + \Delta G_{002}$	[Cro] ²	
					$\Delta G_{220} = \Delta G_{020} + \Delta G_{200}$	[Cro] ²	
					$\Delta G_{202} = \Delta G_{200} + \Delta G_{002}$	[Cro] ²	
					$\Delta G_{120} = \Delta G_{100} + \Delta G_{020}$	[Ci][Cro]	
					$\Delta G_{210} = \Delta G_{200} + \Delta G_{010}$	[Ci][Cro]	
					$\Delta G_{102} = \Delta G_{100} + \Delta G_{002}$	[Ci][Cro]	
					$\Delta G_{201} = \Delta G_{200} + \Delta G_{001}$	[Ci][Cro]	
					$\Delta G_{012} = \Delta G_{010} + \Delta G_{002}$	[Ci][Cro]	
					$\Delta G_{021} = \Delta G_{020} + \Delta G_{001}$	[Ci][Cro]	
					$\Delta G_{111} = \Delta G_{100} + \Delta G_{010} + \Delta G_{001} + \Delta G_{coop}$	[Ci] ³	
					$\Delta G_{222} = \Delta G_{200} + \Delta G_{020} + \Delta G_{002}$	[Cro] ³	
					$\Delta G_{112} = \Delta G_{100} + \Delta G_{010} + \Delta G_{coop} + \Delta G_{002}$	[Ci] ² [Cro]	
					$\Delta G_{121} = \Delta G_{100} + \Delta G_{020} + \Delta G_{001}$	[Ci] ² [Cro]	
					$\Delta G_{211} = \Delta G_{200} + \Delta G_{001} + \Delta G_{010} + \Delta G_{coop}$	[Ci] ² [Cro]	
					$\Delta G_{221} = \Delta G_{200} + \Delta G_{020} + \Delta G_{001}$	[Ci][Cro] ²	
					$\Delta G_{212} = \Delta G_{200} + \Delta G_{010} + \Delta G_{002}$	[Ci][Cro] ²	
					$\Delta G_{122} = \Delta G_{100} + \Delta G_{020} + \Delta G_{002}$	[Ci][Cro] ²	
					$\Delta G_{30} = -10.5$	1	Z_y
					$\Delta G_{03} = -13.5$	1	Z_x
					$\Delta G_{33} = \Delta G_{30} + \Delta G_{03}$	1	Z_x, Z_y
					$\Delta G_{31} = \Delta G_{30} + \Delta G_{001}$	[cI]	Z_y
					$\Delta G_{32} = \Delta G_{30} + \Delta G_{002}$	[Cro]	Z_y
					$\Delta G_{103} = \Delta G_{100} + \Delta G_{03}$	[cI]	Z_x
					$\Delta G_{203} = \Delta G_{200} + \Delta G_{03}$	[Cro]	Z_x
					$\Delta G_{013} = \Delta G_{010} + \Delta G_{03}$	[cI]	Z_x
					$\Delta G_{023} = \Delta G_{020} + \Delta G_{03}$	[Cro]	Z_x
					$\Delta G_{113} = \Delta G_{100} + \Delta G_{010} + \Delta G_{03}$	[cI] ²	Z_x
					$\Delta G_{123} = \Delta G_{100} + \Delta G_{020} + \Delta G_{03}$	[cI][Cro]	Z_x
					$\Delta G_{213} = \Delta G_{200} + \Delta G_{010} + \Delta G_{03}$	[cI][Cro]	Z_x
					$\Delta G_{223} = \Delta G_{200} + \Delta G_{020} + \Delta G_{03}$	[Cro] ²	Z_x

Table 4.2: Values of variables referenced by the partition function with RNA polymerase states included. A.22 uses the values of $[RNAp]$, ΔG_{30} , and ΔG_{03} listed first, whereas A.23 and A.24 use the values listed second.

5 Conclusions and Future Directions

Daß es mir - oder Allen - so scheint, daraus folgt nicht, daß es so ist.
Ludwig Wittgenstein *On Certainty*

We have compared the probability that an unknown scientist should have found out what has been vainly sought for so long, with the probability that there is one madman the more on the Earth, and the latter has appeared to us the greater.
Henri Poincaré *Science and Hypothesis*

The particular contributions of this author towards this thesis, and towards other dynamical systems not discussed in this thesis, are archived in [Borggren \[2011\]](#). This has allowed for much deeper quantitative inspection of the dynamical systems with noise then hitherto discussed with the phage λ . This has allowed for the prediction of behaviour of phenotypes not yet synthesized and further comparison with those that have been. Some subtlety involved with the removal of RNA polymerase from the partition function has been discussed. The λ phage has been discussed in physical and computational terms in hopes to present the system to a physics, mathematics, and computational audience. The λ phage is proving to be to biophysics what hydrogen was for quantum mechanics¹ and as such is a perfect testing ground for empirical and ab initio physical, computational and mathematical methods for gene expression.

Efforts should be made to link the connection with the decision portion of the cell cycle to the bigger picture involving the upstream and downstream reactions and understand the connections between all the parameters and variables, see [St. Pierre \[2009\]](#), [Arkin and Ross \[1998\]](#). Solutions are only initial conditions for the next portion of the cell cycle and our simple models are in danger of being too simple, as we are not left with much room under the rug to sweep degrees of freedom we choose to ignore.

In particular, volume should be connected into the modeling with greater care. The binding and unbinding are in the end very specific bond types occurring at very specific times, and proceeding in terms of idealized concentrations tends to wash out these subtleties. I cannot say with confidence if the model here can or

¹A point Jin Wang has made to motivate this study.

cannot corroborate the findings of [St. Pierre \[2009\]](#) with respect to the sensitivity of the lysis and lysogeny decision being governed by volume. It can be said though, that the decision being sensitive to volume is not mutually exclusive to the decision also being controlled by protein concentrations. In the mathematical sense, we expect the spaces to be connected; volume is in the end the denominator of these densities. Much of the fluctuation can be considered fluctuation of specific volume of cI. For example, if one were to carefully wrap a cylinder around the helix-bend-helix motif of the protein and associated dimerizations and tabulate the volume, that volume would vary greatly over the course of, say, a binding and unbinding process. This model lacks any specific protein configuration and in general, specific volume can vary a great deal over this space. Discrete jumps in that space would appear as noise in this model and would need to be tabulated as such in the diffusion tensors, for example. From a phenomenological perspective it is a useful construct that is parameterized, ideally, by direct measurement and is a highly intuitive way to reproduce the general features of the circuit.

5.0.1 On Computation and Information

It is important to realize that computers evolve simultaneously in two directions. There is the physical interface, the its that make the bits and the its that control them, but there is also the human interface that establishes how we control the its that control the bits, and how the consequent bits are presented back to us in such a way usable to our central nervous system.

5.0.1.1 The Physical Interface

The physical interface evolves in the direction of greater complexity. Moore's law illustrates famously that rates of (serial and electronic) computation are limited by size and heat production. Parallelization, reversible logic, and quantum computation have all been under continual development to circumvent these issues and with considerable success. Where a decade ago one would speak of how fast their computer was, now one speaks of how many cores it has. This is parallelization's contribution. \aleph_0 core would be the limiting case if this trend of the processor to *duo* – core to *quad* – core to 2^n core persists, but appended to the technical difficulties involved with space are now the technical difficulties involved with time, the synchronization of the bits from all the its. The limitation that won't go away here is the speed of light; if A needs a signal from B before time t , the situation is hopeless unless B is spatially located within ct of A. Multiple processors on a chip enable a reduction in heat dissipated by providing channels for which heat can be carried away from a processor. Cells to cells exhibit analogous, although much more sophisticated signaling, are far closer to the \aleph_0 – core limit, and being waterproof would certainly be convenient for transporting heat elsewhere; 98.6 degrees is certainly quite a clue that bodies are quite adept at managing this

Figure 5.1: Variational forms used in this work.

Python code

```
a = v*u*dx + 0.5*k*(inner(grad(v), -velocity*u)*dx +
    inner(grad(v), D*grad(u))*dx)
L = v*u0*dx - 0.5*k*(inner(grad(v), -velocity*u0)*dx +
    inner(grad(v), D*grad(u0))*dx)
```

housekeeping heat. Reversible logic evades the heat issue while exacerbating the spatial issue since intermediate steps are kept.

5.0.1.2 The Human Interface

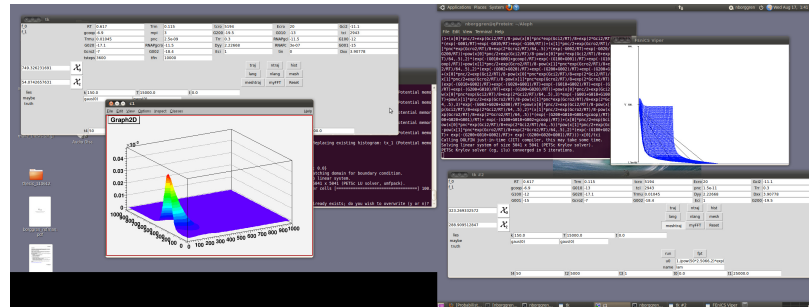
The human interface evolves in the direction of greatest simplicity. Quantum Computer, classical computer, topological quantum computer, chemical computer,...is trivia far from the concern of the user who ultimately just wants to push a button and get a reliable answer. Contrast the Fortran punch codes of old with the following lines of python code that defines our variational forms:

These lines of code are the same no matter what mesh we choose, velocity we choose, time step, initial condition, and diffusion tensor. When the code is called we better have defined what they mean, but until then they are freely transferrable. See more examples in [Logg et al. \[2012\]](#).

A Gui was built to provide buttons to push to generate dynamical trajectories, Langevin trajectories, mesh trajectories (in which an entire domain is propagated in accordance with the flow field), and diffusion trajectories. Systems built are the λ phage, λ phage with RNA polymerase, λ phage approximating O_L states, the cyclin system (see [Wang et al. \[2008\]](#)), the Lorenz attractor, and can operate with imaginary numbers for the ode case.

Here is a screenshot showing multiple instances of the Gui from [Borggren \[2011\]](#).

Figure 5.2: This is the \aleph_{00000} Gui demonstrating user interaction with equations of motion. It reads equations of motion of a system and builds a Gui to interact with that system



5.1 *On the importance of generality and the abstract*

It is important that methods and calculations be *reproducible* and transferrable to different systems. The study of dynamics in the most general settings cannot be neglected for proper understanding of the circuitry and signaling of cells.

Sun Ra says that it is laughable when knowledge is attributed to a man, [Ra and Mugge \[1980\]](#), and this rings all the more true when the discussion is one that seeks to synthesize elements from the great discoveries of physics: quantum mechanics and statistical mechanics, with the great discoveries of biology, evolution, genes, and DNA to just begin. Chemistry has found itself central to the discussions of biology as a result of being best positioned and adept at comprehending the molecular basis and even the purest of mathematics at this stage cannot escape its insights from being applied to the problems of life. At this interface; engineers, medical doctors, biologists, physicists, mathematicians, chemists, logicians, theologians and everybody else with hands and a mind are trying to contribute to this story. The vacuousness of the borders of our disciplines has become evident and we find Nature ready and willing to adapt to any of our biases. The wisdom of Poincaré on mathematics equally applies to the whole of science as well:

We know very well that mathematics will continue to develop, but we have to find out in what direction. I shall be told "in all directions," and that is partly true; but if it were altogether true, it would become somewhat alarming. Our riches would soon become embarrassing, and their accumulation would soon produce a mass just as impenetrable as the unknown truth was to the ignorant.

So it is the case that Science and Method begins with a selection of facts, see [Poincaré \[2010\]](#). For Poincaré this consisted in those with the greatest chance of recurrence and to each of us our lives come to be mirrors of our particular criterion of this selection process. Wittgenstein in his youth would further require that this selection be from all the facts, see [Wittgenstein \[1922\]](#), but by the end of his term he would have brushed that off to some of the symbolic games that we play, see [Wittgenstein \[1975\]](#). The world is more akin to a collection of statements than a collection of facts and a statement can be true, false, or meaningless. Perhaps one may want to continue this line of thinking to get a world that is a collection of words, but words require letters and something to denote nothing to tell where one word ends and the next word begins. The statistical mechanics of random words and strings of letters may be equally important, see [Stephens and Bialek \[2008, 2010\]](#).

What sets the subjects apart on this stage is no longer the content of their subject matter - it is one and the same - but the psychological predispositions of the practitioners. What are the relevant degrees of freedom? No two agree, but nor is it necessary for them to do so. A mutual respect is in order, and as the

tricks of noise and dynamics that can crash markets are now being applied to the problems of life, let us walk gently into this night, none of these days have been promised to us. Recall a giant meteor from outerspace was required to wipe out the dinosaurs and open up our niche. Nobody can look a three-headed pterodactyl with fluorescent wings in the eyes and call themselves the pinnacle of creation, and as far as I can tell, there is nothing prohibiting DNA to make such a genome. Science is still new, our ignorance will be infinite, even while our knowledge can be infinite too. For want of a common goal let us choose, at least, a moral one.

The only thing that could unite the planet is a united space program [health center]... the earth becomes a space station [health center] and war is simply out, irrelevant, flatly insane in a context of research centers, space-ports, and the exhilaration of working with people you like and respect toward an agreed-upon objective, an objective from which all workers will gain. Happiness is a by-product of function. The planetary space station [health center] will give all participants an opportunity to function.

William S. Burroughs, *The Place of Dead Roads* [NAB]

Toss in some clean water, money, and the cure for pain and we may asymptotically approach a civil society. When the subject matter is infection and disease and the consequences are pain and joy it would be imprudent for physics to wait for a topological quantum computer and all orders of precision worked out to chime in to this story, which is still caught up in tabulating everything. From the looks of the White House, and the dazzling noise of stock markets apparently stuck between radio stations, the stalemate seem to persist straight up to the top.

The challenges of this century can provide an incomparable opportunity towards global scientific collaboration as we learn to harness the resources we are enveloped in, or we can watch ourselves fumble the oil into sea, and another mysterious earthquake, tornado and tsunami wipe out our cities, as we choose the right color for our data points, and argue over what data points to delete.

The human genome is at least the magnitude of mystery it was a decade ago and bioinformatics has been more successful at deciphering the code than electricity, magnetism, statistical mechanics, quantum mechanics, and all the other tools physicists pridefully declare universality with. Some basic assumptions clearly need to be relinquished.

5.2 On the importance of the particular and the measurable; Case Studies

A model of the phage λ has been analyzed to illustrate how a particular system is queried. Given that these crystal structures have been measured and good empirical force fields exist, these values of ΔG_s can in principle begin to be predicted and compared to the values deduced from experiments. Ab initio calculations of the O_R region from first principle Hamiltonians may take some time

but until then it is exactly these types of mechanisms that motivate experiments like NSLS to do such crystallographic measurements that empirical protein and DNA force fields are developed to illuminate. Understanding the relation of these nanoscopic structures and dynamics to the macroscopic flows of energy and entropy is the basic question necessary to address in order to appreciate the full value of the accumulated data and begin to use it for clinically relevant purposes. Nuclear magnetic resonance, NMR, is also regularly used to investigate more directly these ensemble properties in vivo, near equilibrium, and can also be used to constrain these numbers. The difference between in vitro and in vivo values for the weights has already been demonstrated to induce topological changes. These numbers have a taxonomy tree of their own as measurement techniques get refined. Our phenomenology here continues to rely on fits to the primary datum from the biological measurements. Physics will learn a lot from the networks that make up biological systems and the manners in which they give rise to emergent degrees of freedom. In order to address the concerns of [Little et al. \[1999\]](#), and perhaps many other scientists who are typically more concerned about large scale qualitative differences, like the difference of lysogeny and lysis, we will scan parameters as much as we can in this volume of spacetime. The particular values of ΔG are dependent in some complicated way to evolution, volume, temperature, pH, doping, in vitro, in vivo and a host of other unknowns, can and do change on large scales. Schrodinger's anthropomorphic remark can be applied here: it is exactly robustness to these types of changes that allow it to exist at 300K for so long in the first place. Here we note a difference in the use of the terminology "robust" between biologists and the use in subsequent theoretical work. For Little, robustness is the property that two states, lytic and lysogenic, are still present after large scale mutations, changes in ΔG due to switching, say, site 3 with a copy of site 1. This is a transformation on the level of the partition function in that the ΔG values change. This differs from the use of robustness in the sense of [Zhu et al. \[2004\]](#), [Wang et al. \[2010\]](#), [Aurell and Sneppen \[2002\]](#), which uses this word to describe the small rate at which a cell state switches spontaneously as a result of noise. The phage can be robust in both senses. It is even robust against our attempts to model it.

Binding studies for drug docking often involve phenomena of this sort and computational drug design as well. This is a largely trial and error process, and perhaps some topological information is due. Biologists are often skeptical of the latest grand unification scheme from the physicists and the applied mathematicians, and their skepticism is mimed here. The main point of separation, as I see it, between the two subjects is in a sense a question of topology. The mass of an electron is a number with small error bars that with hard work can be made even smaller. It is repeatable and exists on other planets too, the numbers of biology are scarce when present. The facts they prefer to select on are topological in nature: to lyse or not to lyse?

5.2.1 Robustness, robustness, stability, and stability

..., I am referring to something that mere words will never be able to express, relative, absolute, full, empty, still alive and no longer alive, because, sir, in case you don't know it, words move, they change from one day to the next, they are as unstable as shadows, are themselves shadows,...

José Saramago, *Death with interruptions*

In the overlap of subjects it is quite expected for there to be ambiguity between the different uses of the same word. This is all too commonplace and has been a perennial source of confusion in this study. Two examples are worthy of note, the notion robust in the title of [Little et al. \[1999\]](#) is different than the use of the same word in the subsequent work of [Wang et al. \[2008\]](#), [Zhu et al. \[2004\]](#). The first use is to denote the propensity of phage λ to maintain a working switch after particular mutations in the O_R sites. This is a change at the level of the partition function; the genome has fundamentally changed, yet the bistability persists, it is robust. The second use refers to something fundamentally different: the capacity of a lysogen to stay a lysogen by continuous maintenance by cI protein to keep, in particular, the cro gene from expression. The use of the word stability in [Little et al. \[1999\]](#) is exactly what [Wang et al. \[2008\]](#), [Zhu et al. \[2004\]](#) refer to as robustness, reserving their use of the word stability to refer to the mathematical use of describing the nature of the stationary points. Stability and robustness are used interchangeably in [Aurell and Sneppen \[2002\]](#).

With a four order magnitude spread in the refereed values of the numbers quantifying robustness, it is possible that there is some misunderstanding here. Indeed the numbers reported as experimental in [Wang et al. \[2010\]](#) are not strictly speaking experimental numbers at all, they involve simulation as well, and Little has now decreased those numbers to 10^{-8} or 10^{-9} , in units of lysogens flipping per generation, but has not published these numbers in a refereed journal and caveats the measurement as very difficult when it arises. Finding a more realistic diffusion tensor and better oligomerization equations of state are likely to help better quantify the robustness in the model. We note that in the large concentration limit, where the dynamics picture is sufficient, the phase space completely partitions into two basins of attraction and the number is simply 0. The stability puzzle is still puzzling.

5.3 Algebraic, Analytical, or Computational Methods?

A computational approach, the finite elements, were eventually decided upon for the practical solution of the differential equations herein [Logg et al. \[2012\]](#). Other methods were available and attempted. In particular, an algebraic approach was also investigated (see [Marcus \[1960\]](#), [Figueiredo et al. \[1998\]](#)), and a brute-force analytic approach was used in [Snider \[1999\]](#). These are briefly discussed but were found not as immediately illuminating as the numerical approach throughout.

5.3.1 Non-Associative Commutative Algebra

A differential equation is associated with a non-associative algebra such that invariant manifolds of the differential equation correspond to invariant sub-algebras of the algebra. In particular, given a dynamical system, expand in a form,

$$\dot{x}_i = \sum_{j=1}^m A_{ij} \prod_{k=1}^n x_k^{B_{jk}}, \quad (5.1)$$

we can then define a matrix $M = BA$ and a system $\dot{U}_i = U_i M_{ij} U_j$. This is then a quadratic system and the methods of Marcus apply; we can define an algebra by introducing a product $u_i * u_j = \frac{1}{2}(\delta_{ik} M_{kj} + \delta_{ij} M_{jk})$.

5.3.2 Recursion Equations for Expansion coefficients

In particular, a dynamical system with some noise is concisely written adopting the summation convention

$$\partial_t P = \partial_i D_{ij} \partial_j P - \partial_i (F_i P), \quad (5.2)$$

Roman indices run from 1 to K over the generalized coordinates of the system. This relates the time change of the surface P with a diffusion term and the flow field of the dynamical system. The noise has effectively been smoothed and encoded in the surface P through the diffusion tensor D_{ij} . To be explicit, by $\partial_i D_{ij} \partial_j P$ we mean $\partial_i (D_{ij} (\partial_j P))$.

The zero is used to denote the time index and thus $K + 1$ indices over the non-negative integers are summed over. Ansatz P , a polynomial expansion of the form

$$P(x_0, x_1, \dots, x_K) = p_{i_0, i_1, \dots, i_K} \prod_{j=0, \dots, K} x_j^{i_j}, \quad (5.3)$$

and similarly for the components of F

$$F_m(x_0, x_1, \dots, x_K) = f_{m, i_0, i_1, \dots, i_K} \prod_{j=0, \dots, K} x_j^{i_j}. \quad (5.4)$$

We have allowed the index corresponding to powers of x_i to run over a countably infinite set if need be, however the m counts only the K indices of the coordinates. In general we need an analogue expansion for D_{mn} . We will soon specify to the constant diffusion case, which is lowest order of the general case

$$D_{mn}(x_0, x_1, \dots, x_K) = D_{m, n, i_0, i_1, \dots, i_K} \prod_{j=0, \dots, K} x_j^{i_j}. \quad (5.5)$$

Term by term we will decipher equation (5.2),

$$\partial_t P(x_0, x_1, \dots, x_K) = p_{i_0, i_1, \dots, i_K} (i_0 x^{i_0-1}) \prod_{j=1, \dots, K} x_j^{i_j} = p_{i_0+1, i_1, \dots, i_K} (i_0 + 1) \prod_{j=0, \dots, K} x_j^{i_j}. \quad (5.6)$$

Here we specify constant diffusion tensors. For the next term in equation (5.2), the coefficients on the diagonal are related by

$$D_{mm} \partial_m \partial_m P(x_0, x_1, \dots, x_K) = D_{mm} p_{i_0, i_1, \dots, i_m+2, \dots, i_K} (i_m + 2)(i_m + 1) \prod_{j=1, \dots, K} x_j^{i_j}, \quad (5.7)$$

and off diagonal by

$$D_{mn} \partial_m \partial_n P(x_0, x_1, \dots, x_K) = p_{i_0, i_1, \dots, i_m+1, \dots, i_n+1, \dots, i_K} (i_m + 1)(i_n + 1) \prod_{j=1, \dots, K} x_j^{i_j}. \quad (5.8)$$

The last term in equation (5.2) has two contributions through the product rule,

$$(\partial_m F_m) P(x_0, x_1, \dots, x_K) = (f_{m, i'_0, i'_1, \dots, i'_m+1, \dots, i'_K} (i'_m + 1) \prod_{j=0, \dots, K} x_j^{i'_j}) p_{i_0, i_1, \dots, i_m, \dots, i_K} \prod_{j=1, \dots, K} x_j^{i_j}, \quad (5.9)$$

$$F_m \partial_m P(x_0, x_1, \dots, x_K) = (f_{m, i'_0, i'_1, \dots, i'_K} \prod_{j=0, \dots, K} x_j^{i'_j}) p_{i_0, i_1, \dots, i_m+1, \dots, i_K} (i_m + 1) \prod_{j=1, \dots, K} x_j^{i_j}. \quad (5.10)$$

These become

$$(\partial_m F_m) P(x_0, x_1, \dots, x_K) = f_{m, i'_0, i'_1, \dots, i'_m+1, \dots, i'_K} (i'_m + 1) p_{i_0, i_1, \dots, i_m, \dots, i_K} \prod_{j=1, \dots, K} x_j^{i'_j + i_j}, \quad (5.11)$$

$$F_m \partial_m P(x_0, x_1, \dots, x_K) = f_{m, i'_0, i'_1, \dots, i'_K} p_{i_0, i_1, \dots, i_m+1, \dots, i_K} (i_m + 1) \prod_{j=1, \dots, K} x_j^{i'_j + i_j}. \quad (5.12)$$

Let us simplify by saying the diffusion tensor is diagonal. Substituting into the differential equation yields

$$\begin{aligned}
p_{i_0+1,i_1,\dots,i_K}(i_0+1) \prod_{j=0,\dots,K} x_j^{i_j} &= D_{mm} p_{i_0,i_1,\dots,i_m+2,\dots,i_K}(i_m+2)(i_m+1) \prod_{j=1,\dots,K} x_j^{i_j} \\
&\quad - f_{m,i'_0,i'_1,\dots,i'_m+1,\dots,i'_K} p_{i''_0,i''_1,\dots,i''_m,\dots,i''_K}(i'_m+1) \prod_{j=1,\dots,K} x_j^{i'_j+i''_j} \\
&\quad - f_{m,i'_0,i'_1,\dots,i'_K} p_{i''_0,i''_1,\dots,i''_m+1'',\dots,i''_K}(i''_m+1) \prod_{j=1,\dots,K} x_j^{i'_j+i''_j}.
\end{aligned} \tag{5.13}$$

Setting the coefficients of $\prod_{j=0,\dots,K} x_j^{i_j}$ equal on both sides gives the recursion relation

$$\begin{aligned}
p_{i_0+1,i_1,\dots,i_K}(i_0+1) &= D_{mm} p_{i_0,i_1,\dots,i_m+2,\dots,i_K}(i_m+2)(i_m+1) \\
&\quad - \sum_{i'_j+i''_j=i_j} f_{m,i'_0,i'_1,\dots,i'_m+1,\dots,i'_K} p_{i''_0,i''_1,\dots,i''_m,\dots,i''_K}(i'_m+1) \\
&\quad - \sum_{i'_j+i''_j=i_j} f_{m,i'_0,i'_1,\dots,i'_K} p_{i''_0,i''_1,\dots,i''_m+1'',\dots,i''_K}(i''_m+1).
\end{aligned} \tag{5.14}$$

A steady state solution in particular requires $p_{i_0} = 0$ for all i . So that the index can be dropped altogether.

$$\begin{aligned}
0 &= D_{mm} p_{i_1,\dots,i_m+2,\dots,i_K}(i_m+2)(i_m+1) \\
&\quad - \sum_{i'_j+i''_j=i_j} f_{m,i'_1,\dots,i'_m+1,\dots,i'_K} p_{i''_1,\dots,i''_m,\dots,i''_K}(i'_m+1) \\
&\quad - \sum_{i'_j+i''_j=i_j} f_{m,i'_1,\dots,i'_K} p_{i''_1,\dots,i''_m+1'',\dots,i''_K}(i''_m+1).
\end{aligned} \tag{5.15}$$

The software of [Borggren \[2011\]](#), can also perform calculations based on these approaches.

5.4 From biology to physics to biology

Physics has always relied heavily on progress from biology for its own progress. The lung predates the steam engine and it is unlikely the latter would have emerged without the scientific inspection of the former. The central nervous system predates the understanding of electricity and again the latter would not have emerged without the scrutiny of the former.

Consider the timeline of the study of metabolism, see [Kilgour \[1955\]](#). In around 1500, Leonardo da Vinci likened the burning of a candle to animal nutrition, from

which he inferred that an atmosphere that could not support combustion would also not support animals. In 1648, Van Helmont coined the word 'gas' in a journal for medicine under the title *Ortus Medicinae*. The first gas described as such was CO_2 , the biproduct of animal respiration. From 1660-1678 Robert Boyle developed experiments that studied the changes in a volume of air due to respiration and combustion and similarly for Mayow's *Treatise on Respiration*. In 1754 Black discovered *fixed air*, CO_2 . In 1766, Cavendish discovered *inflammable air*, H_2 . In 1772, Rutherford described *residual air*, N_2 . In 1780, with their new invention, the *calorimeter*, Lavoisier and Laplace concluded that da Vinci's hypothesis of respiration being a form of combustion is indeed the case in their beautiful paper *Memoir on Heat*. Thus the groundwork had been laid for thermodynamics and the subsequent industrial revolution, the driving force of which having been the study of respiration.

One only need to read the bibliography of Hermann Helmholtz to see that physiology was a driving force in the study of electricity and acoustics, see [Helmholtz \[1954\]](#).

Progress in physics in the 20th century, given the equipment and good mathematical abstractions of the 19th century, managed on its own for a century without much aid from biology and chemistry, but now it is time to revisit the old phenomena, with our new lights, and actually solve some problems.

Biologists and chemists have handed us the structures of their molecules, the wirings of their systems, and the phase diagrams of their compounds. If those are the answers and statistical mechanics and quantum mechanics provide the questions, what is left but the proof?

Bibliography

References

John W. Little, Donald P. Shepley, and David W. Wert. Robustness of a gene regulatory circuit. *EMBO J*, 18(15):4299–4307, 1999.

Mark Ptashne. Principles of a switch. *Nature Chemical Biology*, 7, 2011.

P.W. Anderson. More is different. *Science*, 177(4047):393–396, 1972.

Rolf Landauer. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5(3):183–191, 1961.

Ernst Mayr. *What evolution is*. Basic Books, 2002.

M. Ptashne. *A genetic switch: phage λ and higher organisms*. Wiley-Blackwell, 1992. ISBN 0865422095.

Shota Atsumi and John W. Little. A synthetic phage λ regulatory circuit. *Proceedings of the National Academy of Sciences*, 103(50):19045–19050, December 2006. doi: 10.1073/pnas.0603052103. URL <http://www.pnas.org/content/103/50/19045.abstract>.

Madeline A. Shea and Gary K. Ackers. The OR control system of bacteriophage lambda : A physical-chemical model for gene regulation. *Journal of Molecular Biology*, 181(2):211–230, January 1985. ISSN 0022-2836. doi: 10.1016/0022-2836(85)90086-5. URL <http://www.sciencedirect.com/science/article/B6WK7-4DN3YD2-3T/2/e311a1286e1d6ddf359f9cf6eeea1240>.

Justin B. Kinney, Gašper Tkačik, and Curtis G. Callan. Precise physical models of protein-dna interaction from high-throughput data. *Proceedings of the National Academy of Sciences*, 104(2):501–506, 2007.

Francois Jacob. Evolution and tinkering. *Science*, 1977.

Barry McCoy. *Advanced Statistical Mechanics*. Oxford University Press, 2010.

Charles Darwin. Biodiversity heritage library, 2011. URL <http://www.biodiversitylibrary.org/collection/darwinlibrary>.

Tree of life web project, 2011. URL <http://tolweb.org/tree/>.

Tommaso Toffoli. Physics and computation. *International Journal of Theoretical Physics*, 21(3-4):165–175, April 1982. ISSN 0020-7748. doi: 10.1007/BF01857724. URL <http://www.springerlink.com/content/j88h4652vh227853/>.

S. Lloyd. *Programming the universe*. Knopf, 2006.

C.H. Bennett. The thermodynamics of computation—A review. *International Journal of Theoretical Physics*, 21(12):905–940, 1982.

M. Ptashne and A. Gann. *Genes & signals*. CSHL Press, 2002. ISBN 0879696338.

C. Nüsslein-Volhard. *Coming to life: how genes drive development*. Kales Pr, 2006.

F. St. Pierre. *Determination of Cell Fate Selection During Phage λ Infection*. MIT, 2009.

X-M Zhu, L Yin, L Hood, and P Ao. Robustness, stability and efficiency of phage lambda genetic switch: dynamical structure analysis. *J Bioinform Comput Biol*, 2(4):785–817, 2004. ISSN 0219-7200. URL <http://www.biomedsearch.com/nih/Robustness-stability-efficiency-phage-lambda/15617166.html>.

Jin Wang, Kun Zhang, and Erkwang Wang. Kinetic paths, time scale, and underlying landscapes: a path integral framework to study global natures of nonequilibrium systems and networks. *The Journal of Chemical Physics*, 133(12):125103, September 2010. ISSN 1089-7690. doi: 10.1063/1.3478547. URL <http://www.ncbi.nlm.nih.gov/pubmed/20886967>. PMID: 20886967.

Adam Arkin and John Ross. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected escherichia coli cells. *Genetics*, 149(4):1633–1648, 1998.

Martin Guthold, Xingshu Zhu, Claudio Rivetti, Guoliang Yang, Neil H. Thomson, Sandor Kasas, Helen G. Hansma, Bettye Smith, Paul K. Hansma, and Carlos Bustamante. Direct observation of one-dimensional diffusion and transcription by escherichia coli rna polymerase. *Biophysical Journal*, 77(4):2284 – 2294, 1999. ISSN 0006-3495. doi: 10.1016/S0006-3495(99)77067-0. URL <http://www.sciencedirect.com/science/article/pii/S0006349599770670>.

A. M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):37–72, 1952. ISSN 0080-4622. URL <http://www.jstor.org/stable/92463>. ArticleType: research-article / Full publication date: Aug. 14, 1952 / Copyright © 1952 The Royal Society.

Shigeru Kondo and Takashi Miura. Reaction-Diffusion model as a framework for understanding biological pattern formation. *Science*, 329(5999):1616 –1620, 2010. doi: 10.1126/science.1179047. URL <http://www.sciencemag.org/content/329/5999/1616.abstract>.

Hasso Niemann, Roger V Yelle, Wayne T Kasprzak, Thomas E Cravens, Janet G Luhmann, Ralph L McNutt, Wing-Huen Ip, David Gell, Virginie De La Haye, Ingo MÄijller-Wordag, Brian Magee, Nathan Borggren, Steve Ledvina, Greg Fletcher, Erin Walter, Ryan Miller, Stefan Scherer, Rob Thorpe, Jing Xu, Bruce Block, Ken Arnett, and J Hunter Waite. Ion neutral mass spectrometer results from the first flyby of titan. *Science*, 308(5724):982–986, 2005.

I. C. F. MÄijller-Wodarg, R. V. Yelle, N. Borggren, and J. H. Waite Jr. Waves and horizontal structures in titan’s thermosphere. *Journal of Geophysical Research*, 111: 16 PP., December 2006. doi: 200610.1029/2006JA011961. URL <http://www.agu.org/pubs/crossref/2006/2006JA011961.shtml>.

Jin Wang, Li Xu, and Erkang Wang. Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations. *Proceedings of the National Academy of Sciences*, 105(34):12271–12276, 2008. doi: 10.1073/pnas.0800579105. URL <http://www.pnas.org/content/105/34/12271.abstract>.

A. Liapounoff. Untersuchung eines besonderen falles des problems von der stabilitat der bewegung. *Matematicheskii Sbornik*, 17(2):253–333, 1893.

Vito Volterra. *Lecçons sur la theÅorie mathématique de la lutte pour la vie*., Gauthier-Villars et cie., Paris, 1931.

V.I. Arnold. *Ordinary Differential Equations*. The MIT Press, July 1978. ISBN 0262510189.

Lawrence Perko. *Differential equations and dynamical systems*. Springer, 2001. ISBN 9780387951164.

E. Aurell and K. Sneppen. Epigenetics as a first exit problem. *Physical Review Letters*, 88(4):048101, January 2002. doi: 10.1103/PhysRevLett.88.048101. URL <http://link.aps.org/doi/10.1103/PhysRevLett.88.048101>.

Erwin Schrödinger. *What is life?: the physical aspect of the living cell ; with, Mind and matter ; & Autobiographical sketches*. Cambridge University Press, January 1992. ISBN 9780521427081.

Erik Aurell, Stanley Brown, Johan Johanson, and Kim Sneppen. Stability puzzles in phage lambda. *Physical Review E*, 65(5):051914, May 2002. doi: 10.1103/PhysRevE.65.051914. URL <http://link.aps.org/doi/10.1103/PhysRevE.65.051914>.

J. Rothstein. Physics of selective systems: Computation and biology. *International Journal of Theoretical Physics*, 21(3):327–350, 1982.

P Ao. Potential in stochastic differential equations: novel construction. *Journal of Physics A: Mathematical and General*, 37(3):L25–L30, 2004. ISSN 0305-4470. doi: 10.1088/0305-4470/37/3/L01. URL <http://iopscience.iop.org/0305-4470/37/3/L01>.

A. Lwoff. Nobel Lectures, Physiology or Medicine 1963-1970, 1972.

Hao Ge and Hong Qian. Physical origins of entropy production, free energy dissipation, and their mathematical representations. *Physical Review E*, 81(5): 051133, May 2010. doi: 10.1103/PhysRevE.81.051133. URL <http://link.aps.org/doi/10.1103/PhysRevE.81.051133>.

Hong Qian, Saveez Saffarian, and Elliot L. Elson. Concentration fluctuations in a mesoscopic oscillating chemical reaction system. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16):10376–10381, 2002. doi: 10.1073/pnas.152007599. URL <http://www.pnas.org/content/99/16/10376.abstract>.

Anders Logg, Kent-Andre Mardal, and Garth Wells. *Automated Solution of Differential Equations by the Finite Element Method*. Springer, 2012.

Ping Ao. Borges dilemma, fundamental laws, and systems biology. *Bioinformatics and Biology Insights, Bioinformatics and biology insights*, 2, 2:201, 201, 2008. URL <http://ukpmc.ac.uk/articles/PMC2735955/>.

L.M. Adleman. Molecular computation of solutions to combinatorial problems. *Science*, 266(5187):1021, 1994.

L.M. Adleman. Computing with dna. *Scientific american*, 279(8):34–41, 1998.

R.S. Braich, N. Chelyapov, C. Johnson, P.W.K. Rothmund, and L. Adleman. Solution of a 20-variable 3-sat problem on a dna computer. *Science*, 296(5567):499, 2002.

L V Yakushevich, A V Savin, and L I Manevitch. Nonlinear dynamics of topological solitons in DNA. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 66(1 Pt 2):016614, July 2002. ISSN 1539-3755. URL <http://www.ncbi.nlm.nih.gov/pubmed/12241507>. PMID: 12241507.

K.K. Likharev and V.K. Semenov. RSFQ logic/memory family: a new josephson-junction technology for sub-terahertz-clock-frequency digital systems. *IEEE Transactions on Applied Superconductivity*, 1(1):3–28, 1991. ISSN 10518223. doi: 10.1109/77.80745. URL <http://ieeexplore.ieee.org/Xplore/login.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel4%2F77%2F2654%2F00080745.pdf%3Farnumber%3D80745&authDecision=-203>.

- V.K. Semenov, G.V. Danilov, and D.V. Averin. Negative-inductance SQUID as the basic element of reversible josephson-junction circuits. *IEEE Transactions on Applied Superconductivity*, 13(2):938–943, 2003. ISSN 1051-8223. doi: 10.1109/TASC.2003.814155. URL <http://ieeexplore.ieee.org/Xplore/login.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F77%2F27267%2F01211760.pdf%3Farnumber%3D1211760&authDecision=-203>.
- Jie Ren, Vasili K. Semenov, Yuri A. Polyakov, Dmitri V. Averin, and Jaw-Shen Tsai. Progress towards reversible computing <newline/>With nSQUID arrays. *IEEE Transactions on Applied Superconductivity*, 19(3):961–967, 2009. ISSN 1051-8223. doi: 10.1109/TASC.2009.2018250. URL <http://ieeexplore.ieee.org/Xplore/login.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F77%2F5166641%2F05153026.pdf%3Farnumber%3D5153026&authDecision=-203>.
- Jie Ren. *Physically and Logically Reversible Circuit*. Stony Brook University, 2011.
- K. K. Likharev. Classical and quantum limitations on energy consumption in computation. *International Journal of Theoretical Physics*, 21(3):311–326, April 1982. doi: 10.1007/BF01857733. URL <http://dx.doi.org/10.1007/BF01857733>.
- Dmitri V. Averin, Kristian Rabenstein, and Vasili K. Semenov. Rapid ballistic readout for flux qubits. *Physical Review B*, 73(9):094504, March 2006. doi: 10.1103/PhysRevB.73.094504. URL <http://link.aps.org/doi/10.1103/PhysRevB.73.094504>.
- R. Rajaraman. *Solitons and instantons: an introduction to solitons and instantons in quantum field theory*. North-Holland personal library. North-Holland, 1987. ISBN 9780444870476. URL <http://books.google.com/books?id=SGtKSgAACAAJ>.
- Root; a data analysis framework, 2011. URL <http://root.cern.ch/drupal/>.
- Nathan Borggren. Aleph, 2011. URL <https://code.launchpad.net/~nborggren/+junk/Aleph>.
- Philip W. Anderson. Brainwashed by feynman? *Physics Today*, 53(2):11, 2000. ISSN 00319228. doi: 10.1063/1.882955. URL <http://adsabs.harvard.edu/abs/2000PhT....53b..11A>.
- Konstantin K. Likharev. *Dynamics of Josephson junctions and circuits*. CRC Press, 1996. ISBN 9782881240423.
- G K Ackers, A D Johnson, and M A Shea. Quantitative model for gene regulation by lambda phage repressor. *Proceedings of the National Academy of Sciences*, 79(4):1129–1133, February 1982. URL <http://www.pnas.org/content/79/4/1129.abstract>.

Haidong Feng and Jin Wang. Potential and flux decomposition for dynamical systems and Non-Equilibrium thermodynamics: Curvature, gauge field and generalized Fluctuation-Dissipation theorem. *arXiv:1108.5680*, August 2011. URL <http://arxiv.org/abs/1108.5680>.

John Reinitz and J. Rimas Vaisnys. Theoretical and experimental analysis of the phage lambda genetic switch implies missing levels of co-operativity. *Journal of Theoretical Biology*, 145(3):295–318, August 1990. ISSN 0022-5193. doi: 10.1016/S0022-5193(05)80111-0. URL <http://www.sciencedirect.com/science/article/B6WMD-4KD4YDH-2/2/c70a2502f12fcfe1132e3022cc259ff7>.

Jin Wang. Chapter 1 - potential landscape and flux framework of nonequilibrium biological networks. In *Annual Reports in Computational Chemistry*, volume Volume 7, pages 3–37. Elsevier, 2011. ISBN 1574-1400. URL <http://www.sciencedirect.com/science/article/pii/B9780444538352000018>.

Chulan Kwon, Ping Ao, and David J Thouless. Structure of stochastic dynamics near fixed points. *Proceedings of the National Academy of Sciences of the United States of America*, 102(37):13029–13033, September 2005. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0506347102. URL <http://www.pnas.org/content/102/37/13029>.

Chulan Kwon and Ping Ao. Nonequilibrium steady state of a stochastic system driven by a nonlinear drift force. *Physical Review E*, 84(6), December 2011. ISSN 1539-3755, 1550-2376. doi: 10.1103/PhysRevE.84.061106. URL <http://adsabs.harvard.edu/abs/2011arXiv1107.5864K>.

Hong Qian. Vector field formalism and analysis for a class of thermal ratchets. *Physical Review Letters*, 81(15):3063–3066, October 1998. doi: 10.1103/PhysRevLett.81.3063. URL <http://link.aps.org/doi/10.1103/PhysRevLett.81.3063>.

Michael S Samoilov and Adam P Arkin. Deviant effects in molecular reaction pathways. *Nature Biotechnology*, 24(10):1235–1240, October 2006. ISSN 1087-0156. doi: 10.1038/nbt1253. URL <http://www.nature.com/nbt/journal/v24/n10/full/nbt1253.html>.

Karl Friston and Ping Ao. Free energy, value, and attractors. *Computational and Mathematical Methods in Medicine*, 2012:1–27, 2012. ISSN 1748-670X, 1748-6718. doi: 10.1155/2012/937860. URL <http://www.hindawi.com/journals/cmmm/2012/937860/>.

Jin Wang and Peter Wolynes. Instantons and the fluctuating path description of reactions in complex environments. *The Journal of Physical Chemistry*, 100(4): 1129–1136, January 1996. doi: 10.1021/jp951325i. URL <http://dx.doi.org/10.1021/jp951325i>.

Jin Wang, Bo Huang, Xuefeng Xia, and Zhirong Sun. Funneled landscape leads to robustness of cellular networks: MAPK signal transduction. *Biophysical Journal*, 91(5):L54–L56, September 2006. ISSN 00063495. doi: 10.1529/biophysj.106.086777. URL [http://www.cell.com/biophysj/abstract/S0006-3495\(06\)71873-2](http://www.cell.com/biophysj/abstract/S0006-3495(06)71873-2).

Gerhard Stock, Kingshuk Ghosh, and Ken A. Dill. Maximum caliber: A variational approach applied to two-state dynamics. *Journal of Chemical Physics*, 128(19), May 2008. ISSN 0021-9606. doi: 10.1063/1.2918345. PMID: 18500851 PMCID: 2671656.

Kingshuk Ghosh, Ken A. Dill, Mandar M. Inamdar, Effrosyni Seitaridou, and Rob Phillips. Teaching the principles of statistical dynamics. *American Journal of Physics*, 74(2):123–133, February 2006. doi: 10.1119/1.2142789. URL <http://link.aip.org/link/?AJP/74/123/1>.

Udo Seifert. Entropy production along a stochastic trajectory and an integral fluctuation theorem. *Physical Review Letters*, 95(040602), 2005.

Stefan Klumpp and Terence Hwa. Growth-Rate-Dependent partitioning of RNA polymerases in bacteria. *Proceedings of the National Academy of Sciences*, 105(51):20245–20250, December 2008. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0804953105. URL <http://www.pnas.org/content/105/51/20245>.

Sun Ra and Robert Mugge. A joyful noise, 1980.

Henri Poincaré. *Science and Method*. Cosimo, Inc., July 2010. ISBN 9781616402549.

Ludwig Wittgenstein. *Tractatus logico-philosophicus*. Harcourt, Brace & company, inc., 1922.

Ludwig Wittgenstein. *On certainty*. Wiley-Blackwell, October 1975. ISBN 9780631169406.

Greg J. Stephens and William Bialek. Towards a statistical mechanics of four letter words. *arXiv:q-bio/0801.0253v1*, 2008.

Greg J. Stephens and William Bialek. Statistical mechanics of letters in words. *Physical Review E*, 81, 2010.

Lawrence Marcus. Quadratic differential equations and non-associative algebras. *Contributions to the Theory of Nonlinear Oscillations*, 1960.

A. Figueiredo, T. M. Rocha Filho, and L. Brenig. Algebraic structures and invariant manifolds of differential systems. *Journal of Mathematical Physics*, 39(5):2929, 1998. ISSN 00222488. doi: 10.1063/1.532429. URL <http://link.aip.org/link/JMAPAQ/v39/i5/p2929/s1&Agg=doi>.

A.D. Snider. *Partial differential equations: sources and solutions*. Prentice Hall, 1999. ISBN 9780136743590. URL <http://books.google.com/books?id=19QnAQAAIAAJ>.

Frederick G. Kilgour. Great experiments in biology. *The Yale Journal of Biology and Medicine*, 28(3-4):497–497, December 1955. ISSN 0044-0086. undefinedPMCID: 2603405.

Hermann Helmholtz. *On the Sensations of Tone*. Courier Dover Publications, June 1954. ISBN 9780486607535.

A Appendix; Tabulation of Results

Ineluctable modality of the visible: at least that if no more, thought through my eyes. Signatures of all things I am here to read, seaspawn and seawrack, the nearing tide, that rusty boot. Snotgreen, bluesilver, rust: coloured signs. Limits of the diaphane. But he adds: in bodies. Then he was aware of them bodies before of them coloured. How? By knocking his scone against them, sure. Go easy. Bald he was and a millionaire, maestro di color che sanno. Limit of the diaphane in. Why in? Diaphane, adiaphane. If you can put your five fingers through it, it is a gate, if not a door. Shut your eyes and see.

James Joyce, *Ulysses*

We live, I regret to say, in an age of surfaces.

Oscar Wilde, *The importance of being earnest*

A.1 Results

The phase portraits for the 27 mutations are shown in Tables [A.1](#), [A.2](#) and [A.3](#).

The switch component of 3 of the 9 mutations in Table [A.1](#) are destabilized: λ_{112} , λ_{122} , and λ_{132} . Transients however are still largely important for the out of equilibrium switch properties. 3 of the 9 mutations in Table [A.2](#) are destabilized: λ_{212} , λ_{221} , and λ_{231} . Transients however are still largely important for the out of equilibrium switch properties. 5 of the 9 mutations in Table [A.3](#) are destabilized: λ_{311} , λ_{321} , λ_{322} , λ_{331} and λ_{332} . Transients however are still largely important for the out of equilibrium switch properties.

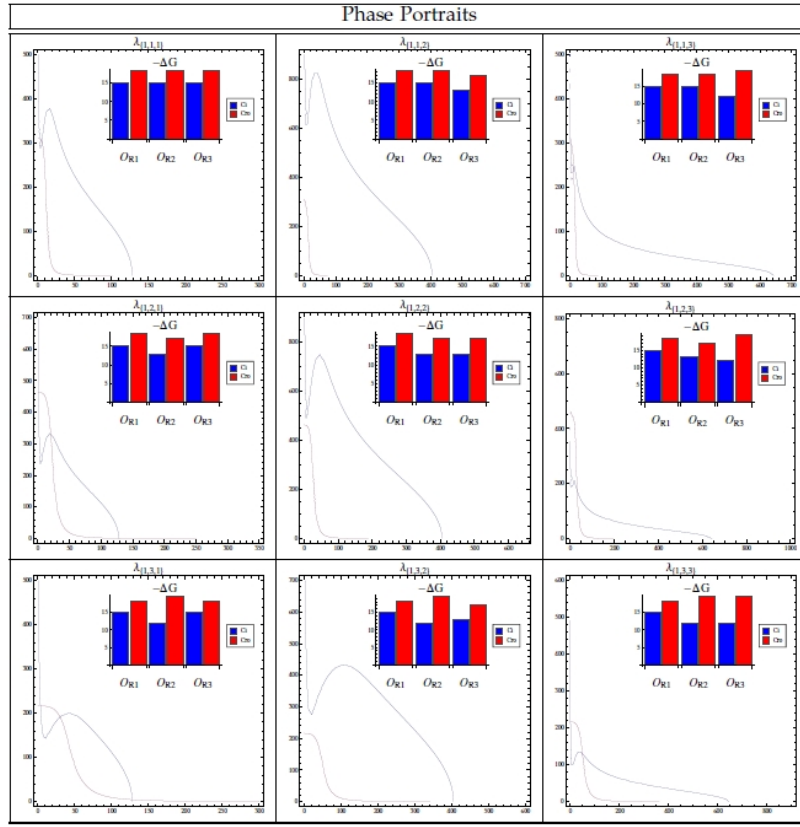
The locations of the stationary points are tabulated in Tables [A.1](#), [A.2](#), [A.3](#), and [A.4](#).

Time evolution of distributions is given in Tables [A.5](#), [A.6](#), [A.7](#) and [A.8](#), corresponding to the wild-type and mutants λ_{121} , λ_{323} , and $\lambda_{323'}$.

The Tables [A.9](#), [A.10](#), [A.11](#), [A.12](#), [A.13](#), [A.14](#), [A.15](#), [A.16](#), [A.17](#), [A.18](#), [A.19](#), [A.20](#), and [A.21](#) archive the results of mutants with three stationary points. Depicted are the phase space portraits alongside the Probability and Flux Landscapes and above the passage time distributions to each of the three stationary points. Assorted expectation values are calculated and can be compared with the wild-type.

Representative trajectories, encoding global features, are depicted in Tables [A.25](#), [A.26](#), and [A.27](#).

Figure A.1: **Contours with stationary Ci and Cro production with O_{R1} fixed. The axis are in protein numbers. 3 of the 9 mutations are destabilized: $\lambda_{112}, \lambda_{122},$ and λ_{132}**



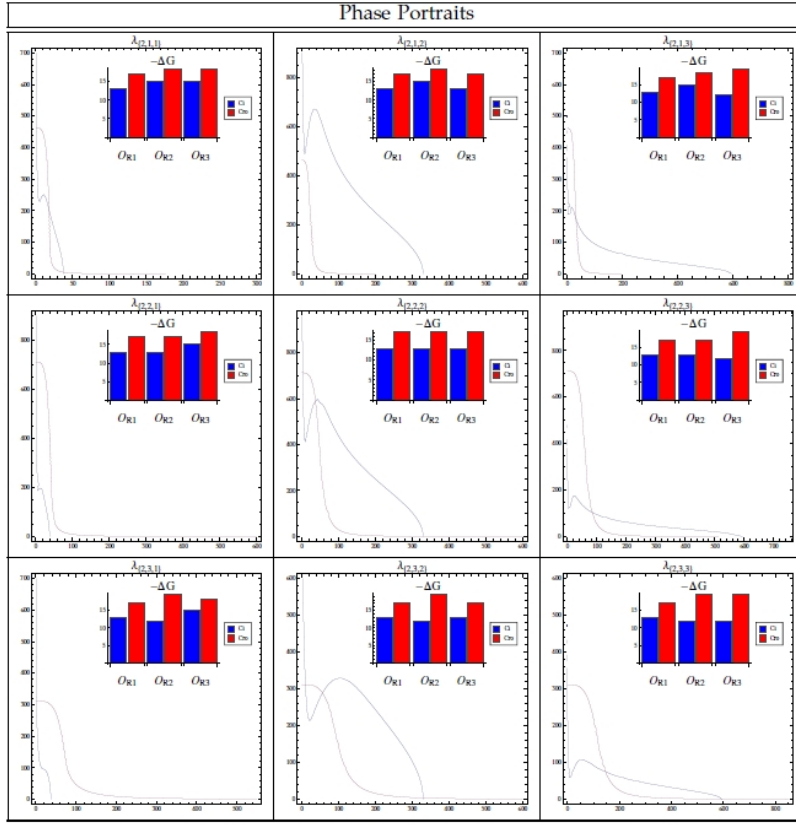
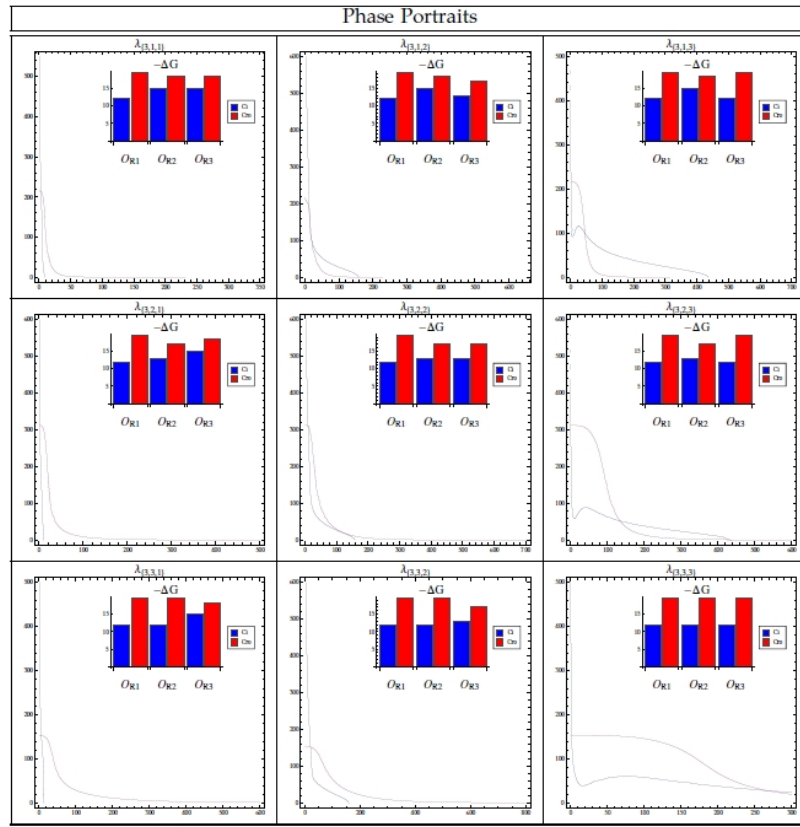


Figure A.2: Contours with stationary Ci and Cro production with O_{R1} fixed. The axis are in protein numbers. 3 of the 9 mutations are destabilized: λ_{212} , λ_{221} , and λ_{231} .

Figure A.3: **Contours with stationary Ci and Cro production with O_{R1} mutated to O_{R3} . 5 of the 9 mutations are destabilized: $\lambda_{311}, \lambda_{321}, \lambda_{322}, \lambda_{331}$ and λ_{332} .**



Properties of the Stationary Point		
Mutant	x (Ci proteins)	y (Cro proteins)
λ_{112}		
	404.90	0.0009
λ_{122}		
	404.89	0.0232
λ_{132}		
	404.89	0.1173
λ_{212}		
	329.88	0.02954
λ_{221}		
	0.3535	712.24
λ_{231}		
	1.7522	312.15
λ_{311}		
	3.4415	214.70
λ_{321}		
	1.7532	312.14
λ_{322}		
	10.253	309.57
λ_{331}		
	6.1379	153.09
λ_{332}		
	20.403	152.35

Table A.1: Mutants with one stationary point.

Properties of Stationary Points			
Mutant	index	x (Ci Proteins)	y (Cro Proteins)
λ_{113}			
	0	644.05	0.00023
	1	0.31299	310.32500

Table A.2: Mutants with two stationary points.

Properties of Stationary Points			
Mutant	index	x (Ci Proteins)	y (Cro Proteins)
λ_{111}			
	0	5.6833	301.287
	1	1.8180	310.218
	2	127.87	0.03220
λ_{121}			
	0	19.9875	332.348
	1	0.82427	462.571
	2	127.86	0.8235
λ_{123} wild-type			
	0	0.141753	462.57273
	1	644.05	0.00611
	2	29.979	178.229
λ_{131}			
	0	30.0039	191.3784
	1	3.43968	216.27658
	2	127.812	4.168042
λ_{133}			
	0	48.6464	130.6941
	1	0.6369	216.311
	2	644.043	0.0309
λ_{211}			
	0	17.505	206.876
	1	37.3015	4.8364
	2	0.8244	462.570
λ_{213}			
	0	0.14175	462.572
	1	595.05	0.00679
	2	29.938	177.711
λ_{222}			
	0	37.234	590.517
	1	2.6864	712.240
	2	329.884	0.7555

Table A.3: Mutants with three stationary points.

Stationary Points			
Mutant	index	x (Ci Proteins)	y (Cro Proteins)
λ_{223}			
	0	0.06002	712.247
	1	595.04	0.17382
	2	83.624	103.62
λ_{232}			
	0	56.762	289.227
	1	10.272	312.072
	2	329.829	3.82207
λ_{233}			
	0	0.30934	312.153
	1	146.0568	74.57208
	2	594.8508	0.87984
λ_{312}			
	0	19.08756	102.4368
	1	13.3744	163.4361
	2	158.8459	0.47835
λ_{313}			
	0	44.4330	100.41068
	1	0.63744	216.3115
	2	436.73205	0.054279
λ_{323}			
	0	0.30935	312.15312
	1	133.92	52.23903
	2	436.115	1.393207
λ_{333}			
	0	1.24472	153.1449
	1	267.603	28.44186
	2	417.693	7.90558

Table A.4: Mutants with three stationary points (continued)

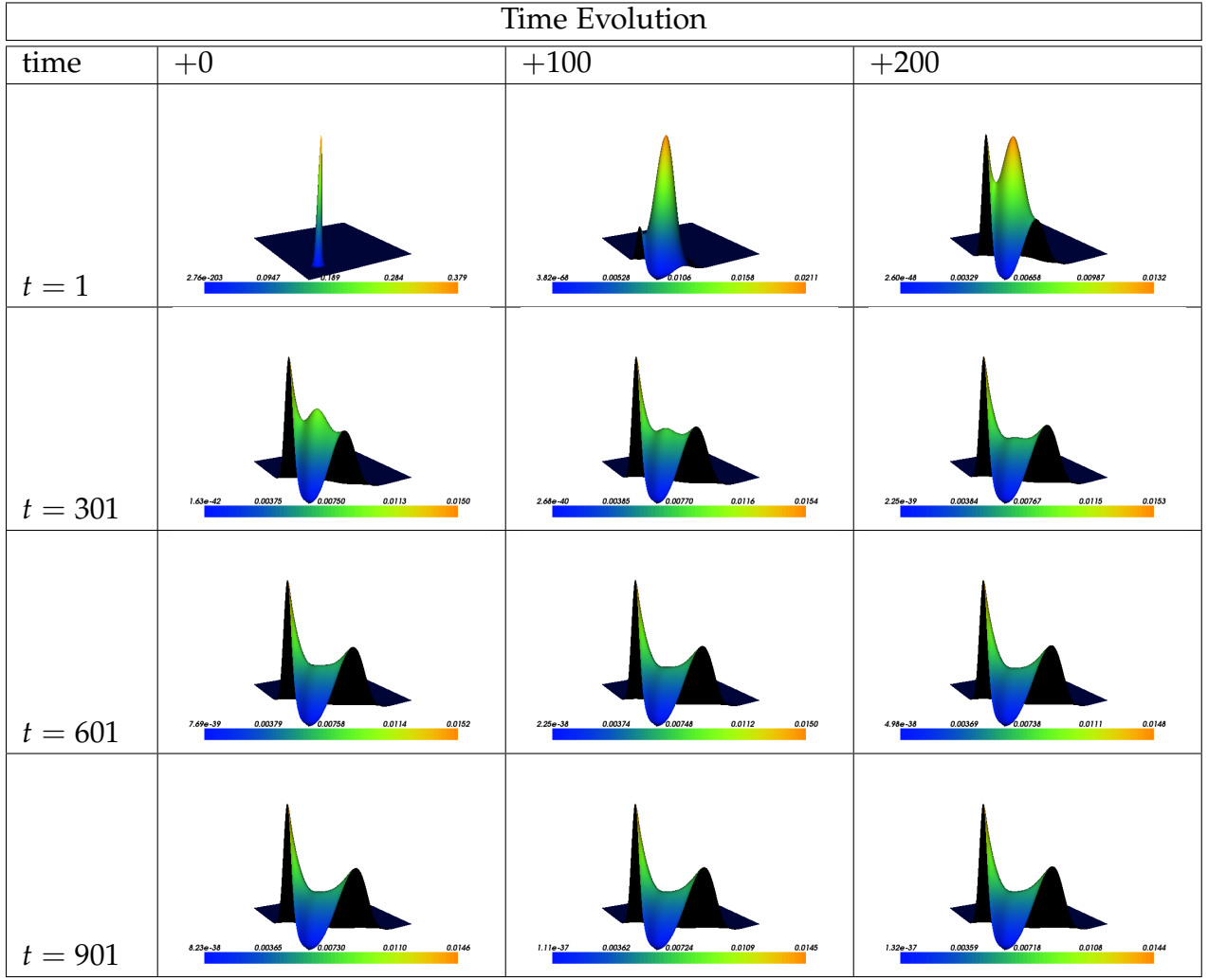


Table A.5: The time evolution of a gaussian initial condition relaxing into the steady state for the wild type. The higher peak corresponds to lytic while the wider peak is lysogenic.

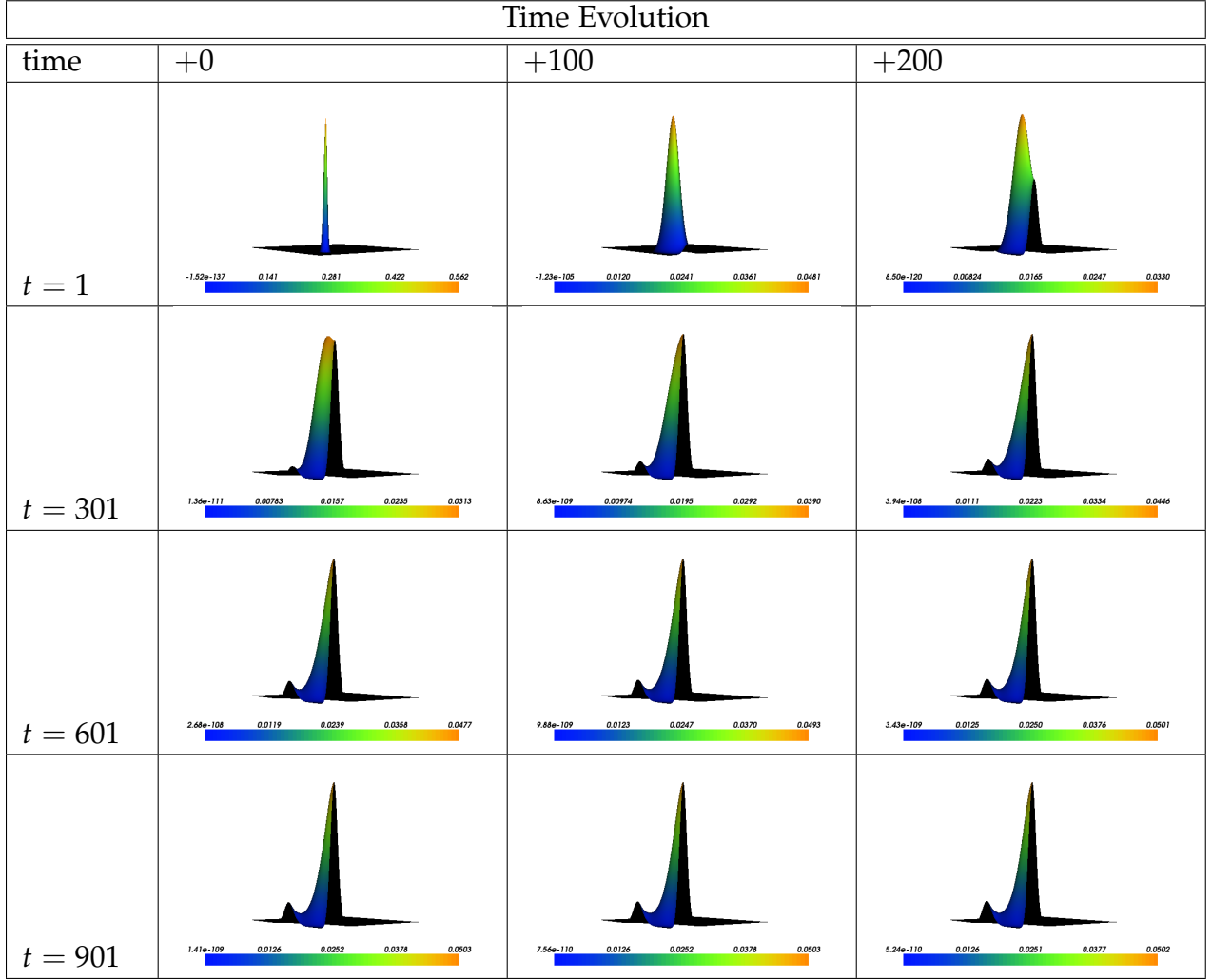


Table A.6: The time evolution of a gaussian initial condition relaxing into the steady state for the 121 mutant. The dominant peak is the lysogenic peak.

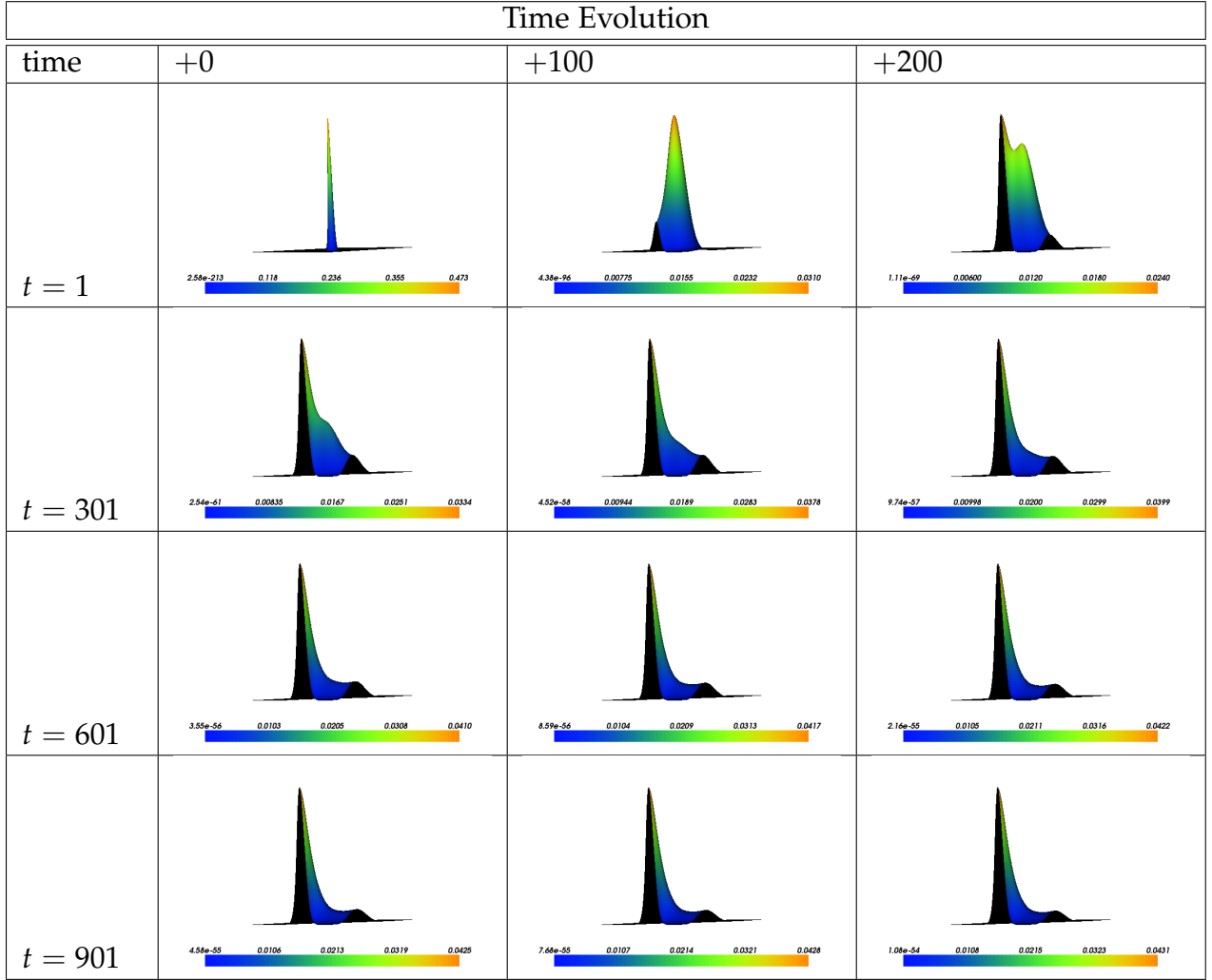


Table A.7: The time evolution of a gaussian initial condition relaxing into the steady state for the 323 mutant. The dominant peak is lytic. The gap is reduce compared to wild type.

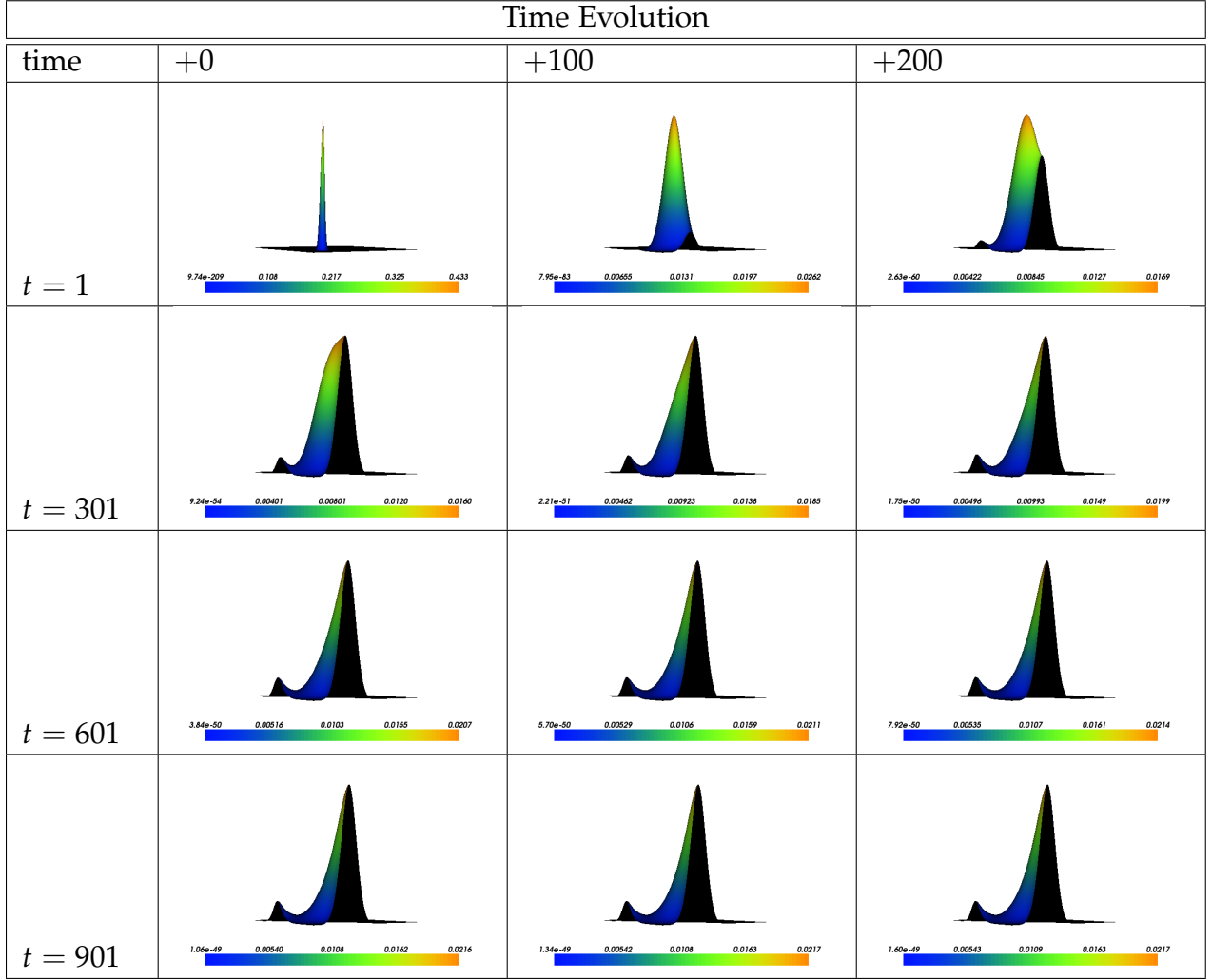
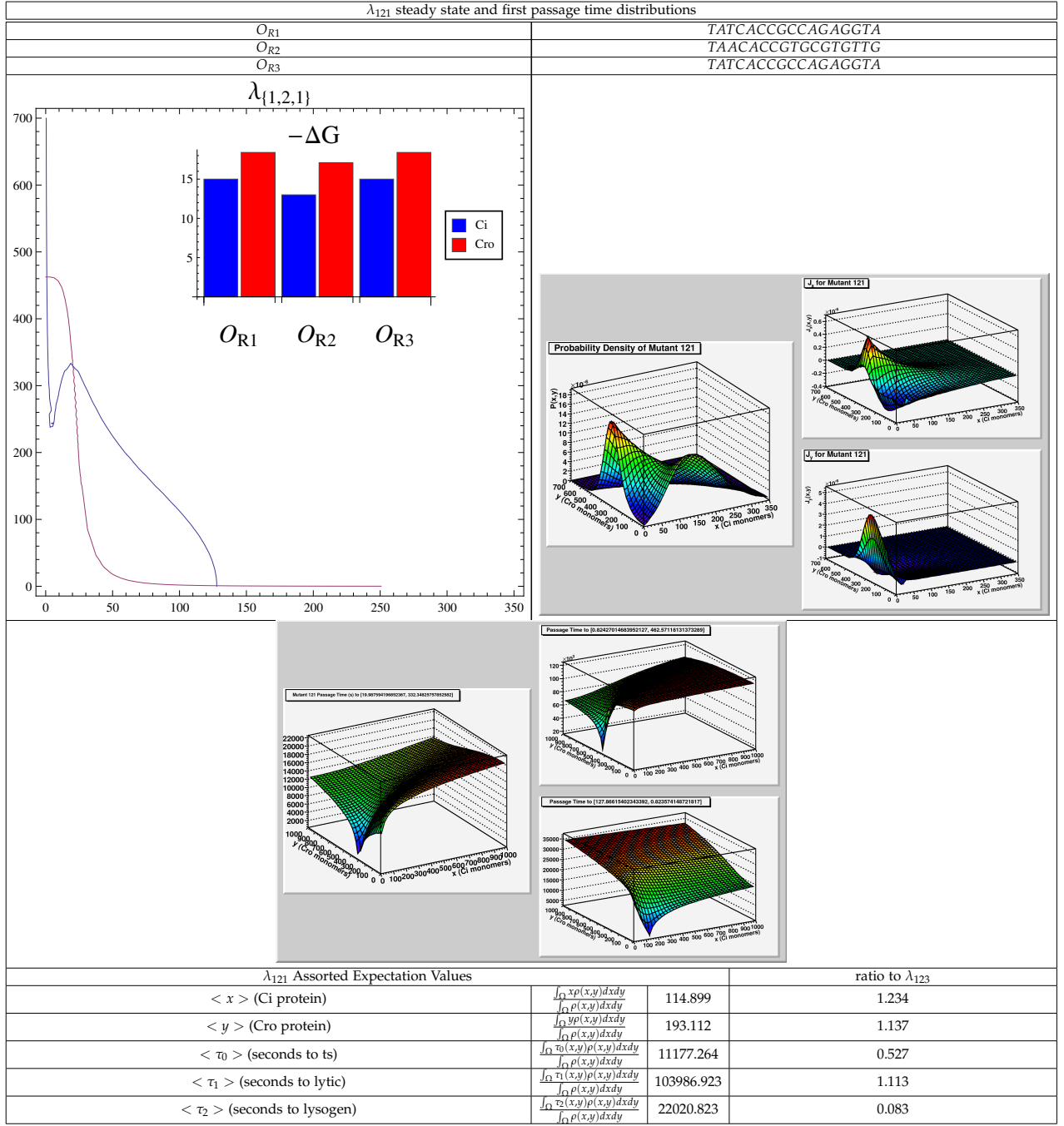
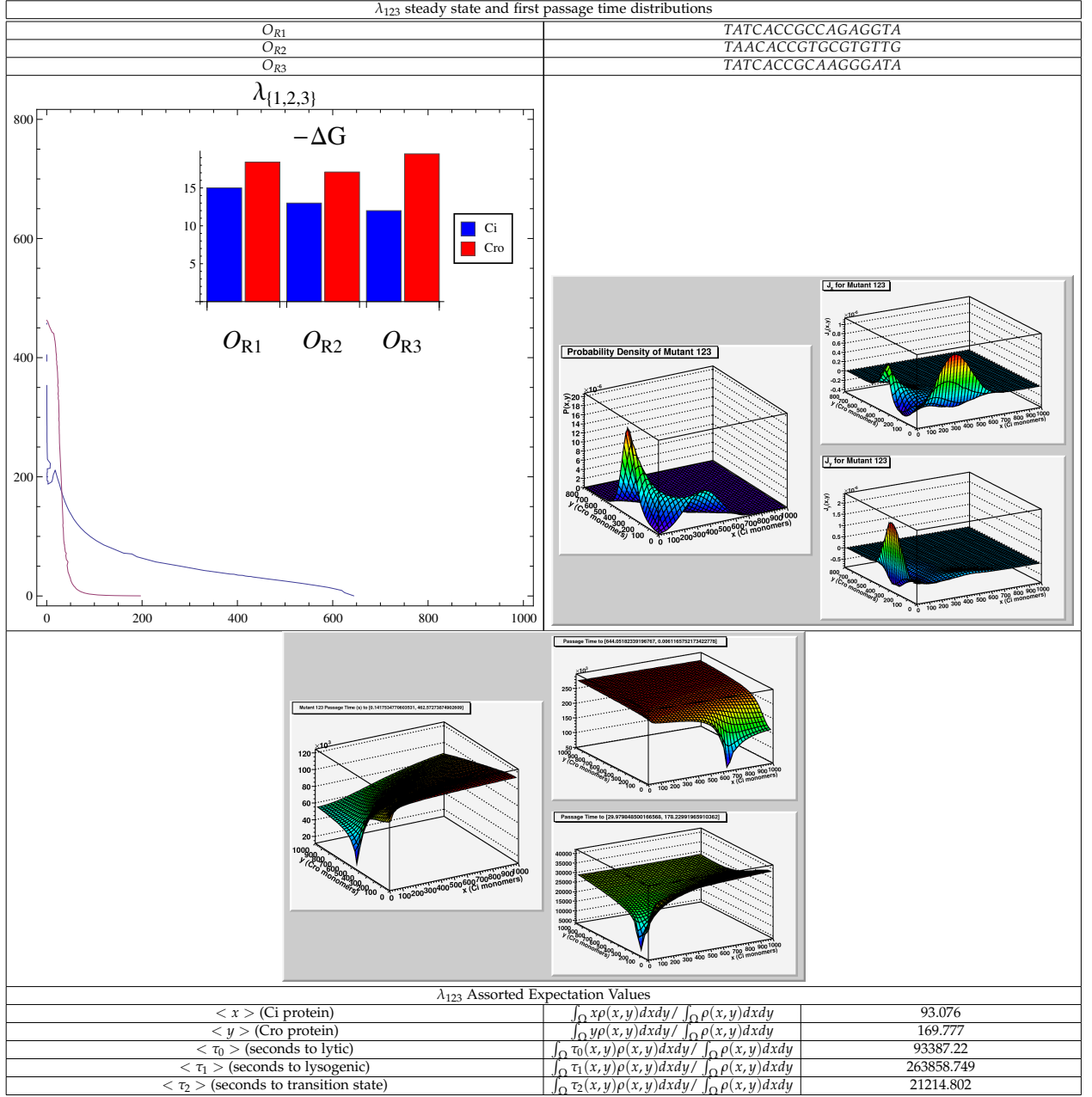
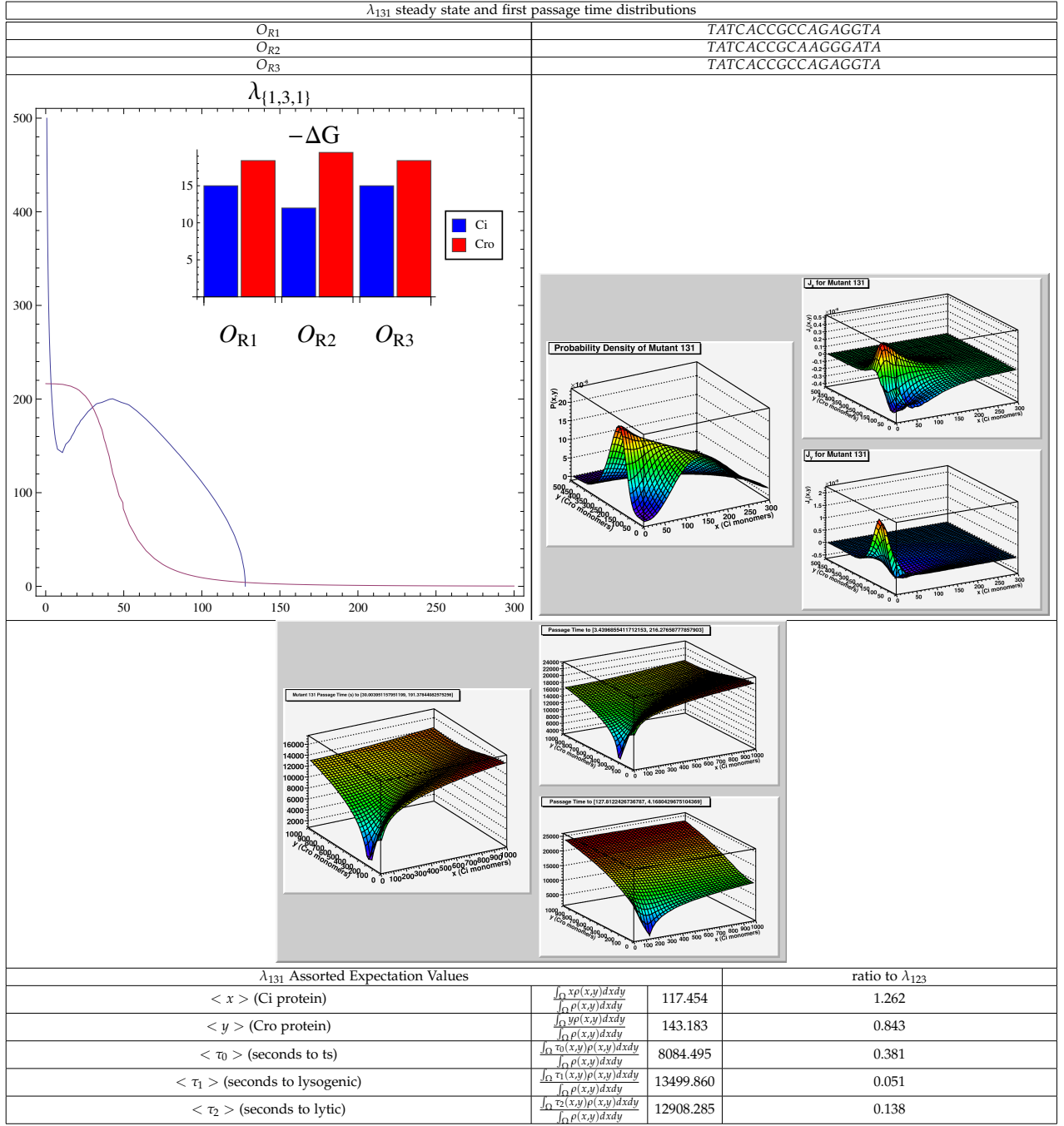


Table A.8: The time evolution of a gaussian initial condition relaxing into the steady state for the 323p mutant. The dominant peak is lysogenic.

Table A.9: Properties of mutant λ_{121}

Table A.10: Properties of wild-type λ_{123}

Table A.11: Properties of mutant λ_{131}

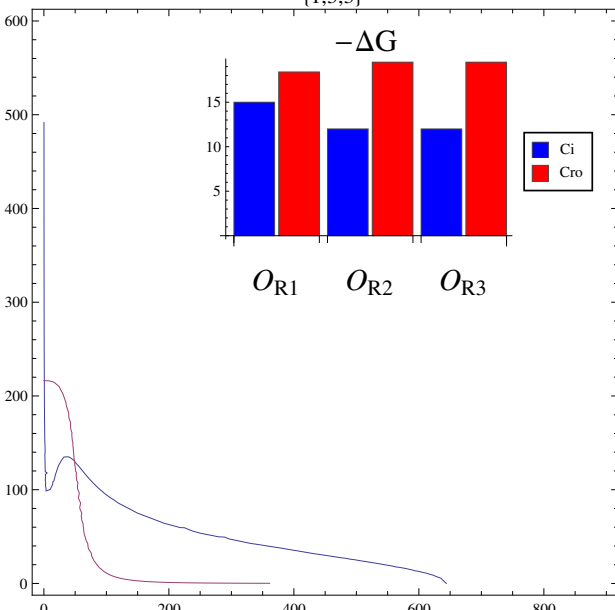
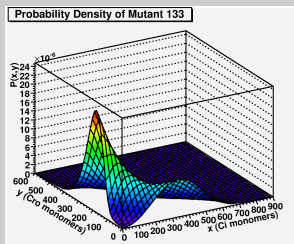
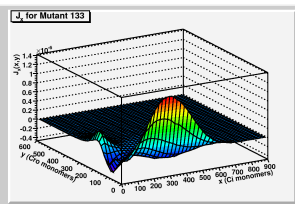
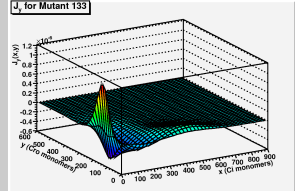
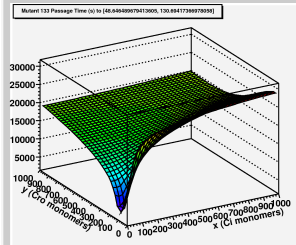
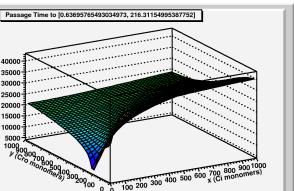
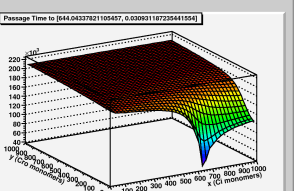
λ_{133} steady state and first passage time distributions			
O_{R1}		TATCACC GCCAGAGGTA	
O_{R2}		TATCACC GCAAGGGATA	
O_{R3}		TATCACC GCAAGGGATA	
<div><div><div><div>$\lambda_{\{1,3,3\}}$</div></div><div><div><div>Probability Density of Mutant 133</div></div><div><div>J_1 for Mutant 133</div></div><div><div>J_2 for Mutant 133</div></div></div></div></div>			
<div><div><div><div>Mutant 133 Passage Time (s) to [0,8269576482634873, 216,31154895387792]</div></div><div><div>Passage Time to [0,8269576482634873, 216,31154895387792]</div></div><div><div>Passage Time to [844,84337821105457, 0,00991187235411554]</div></div></div></div>			
λ_{133} Assorted Expectation Values		ratio to λ_{123}	
$\langle x \rangle$ (Ci protein)	$\frac{\int_{\Omega} x p(x,y) dx dy}{\int_{\Omega} p(x,y) dx dy}$	93.227	1.002
$\langle y \rangle$ (Cro protein)	$\frac{\int_{\Omega} y p(x,y) dx dy}{\int_{\Omega} p(x,y) dx dy}$	137.624	0.8106
$\langle \tau_0 \rangle$ (seconds to ts)	$\frac{\int_{\Omega} \tau_0(x,y) p(x,y) dx dy}{\int_{\Omega} p(x,y) dx dy}$	12264.638	0.578
$\langle \tau_1 \rangle$ (seconds to lytic)	$\frac{\int_{\Omega} \tau_1(x,y) p(x,y) dx dy}{\int_{\Omega} p(x,y) dx dy}$	21628.286	0.232
$\langle \tau_2 \rangle$ (seconds to lysogenic)	$\frac{\int_{\Omega} \tau_2(x,y) p(x,y) dx dy}{\int_{\Omega} p(x,y) dx dy}$	198518.768	0.752

Table A.12: Properties of mutant λ_{133}

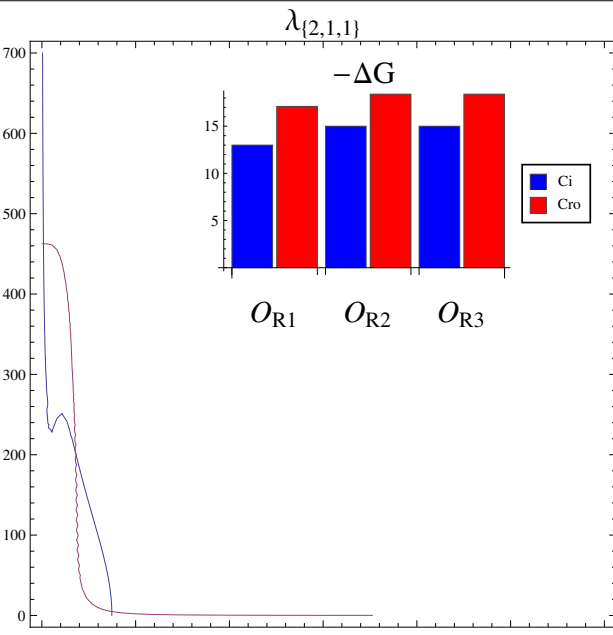
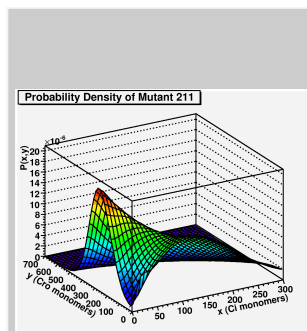
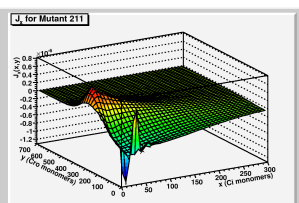
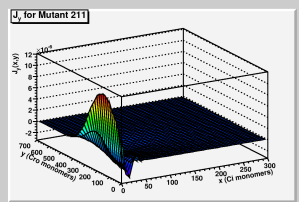
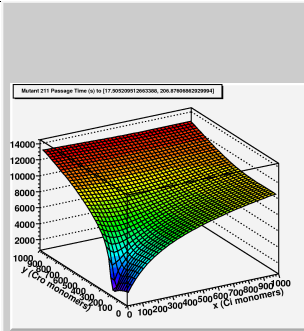
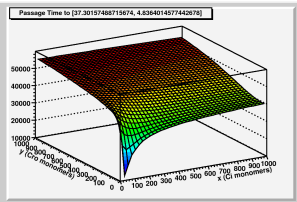
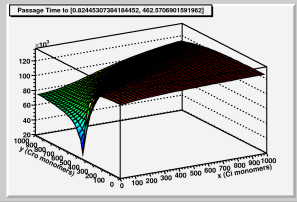
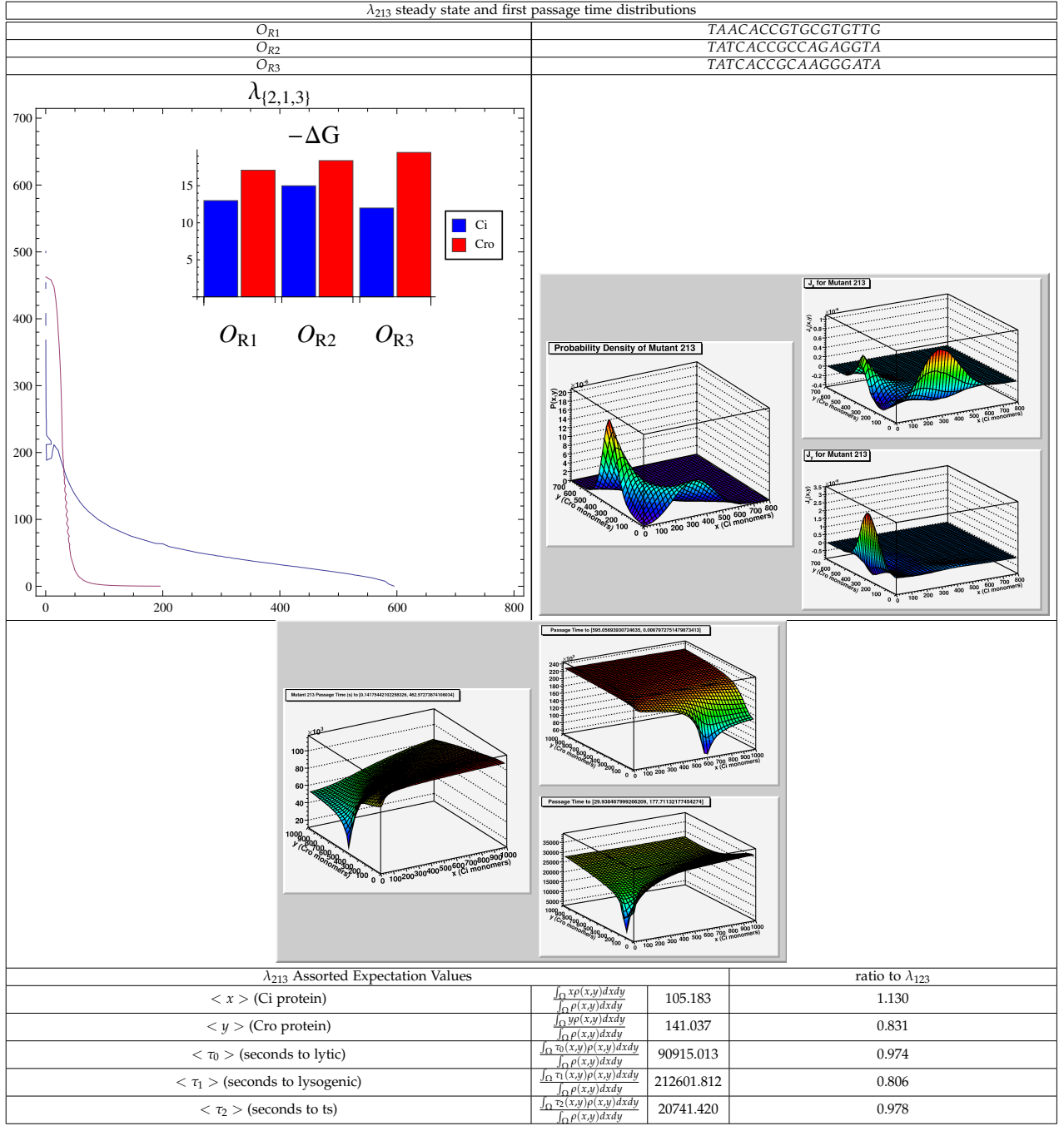
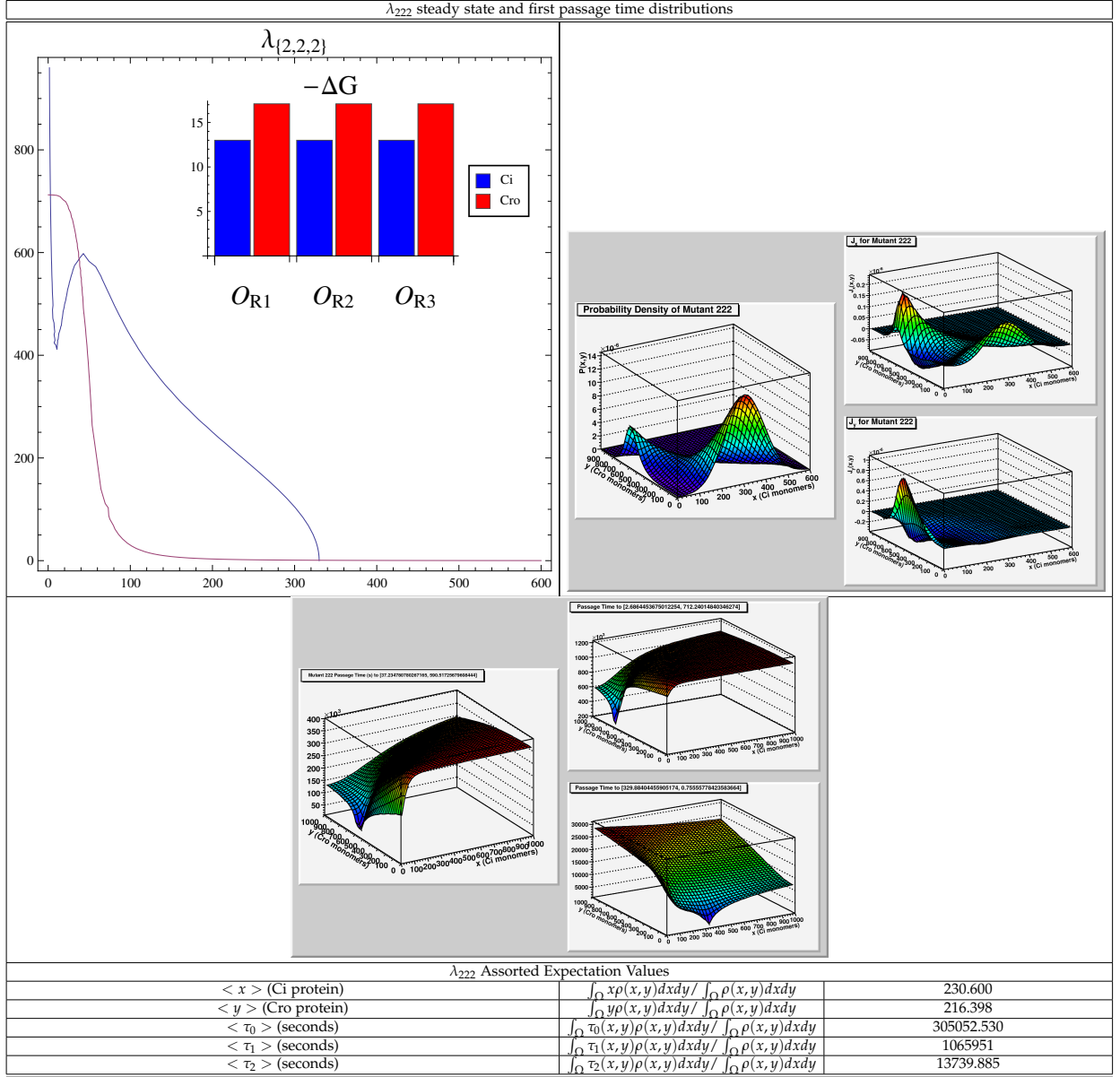
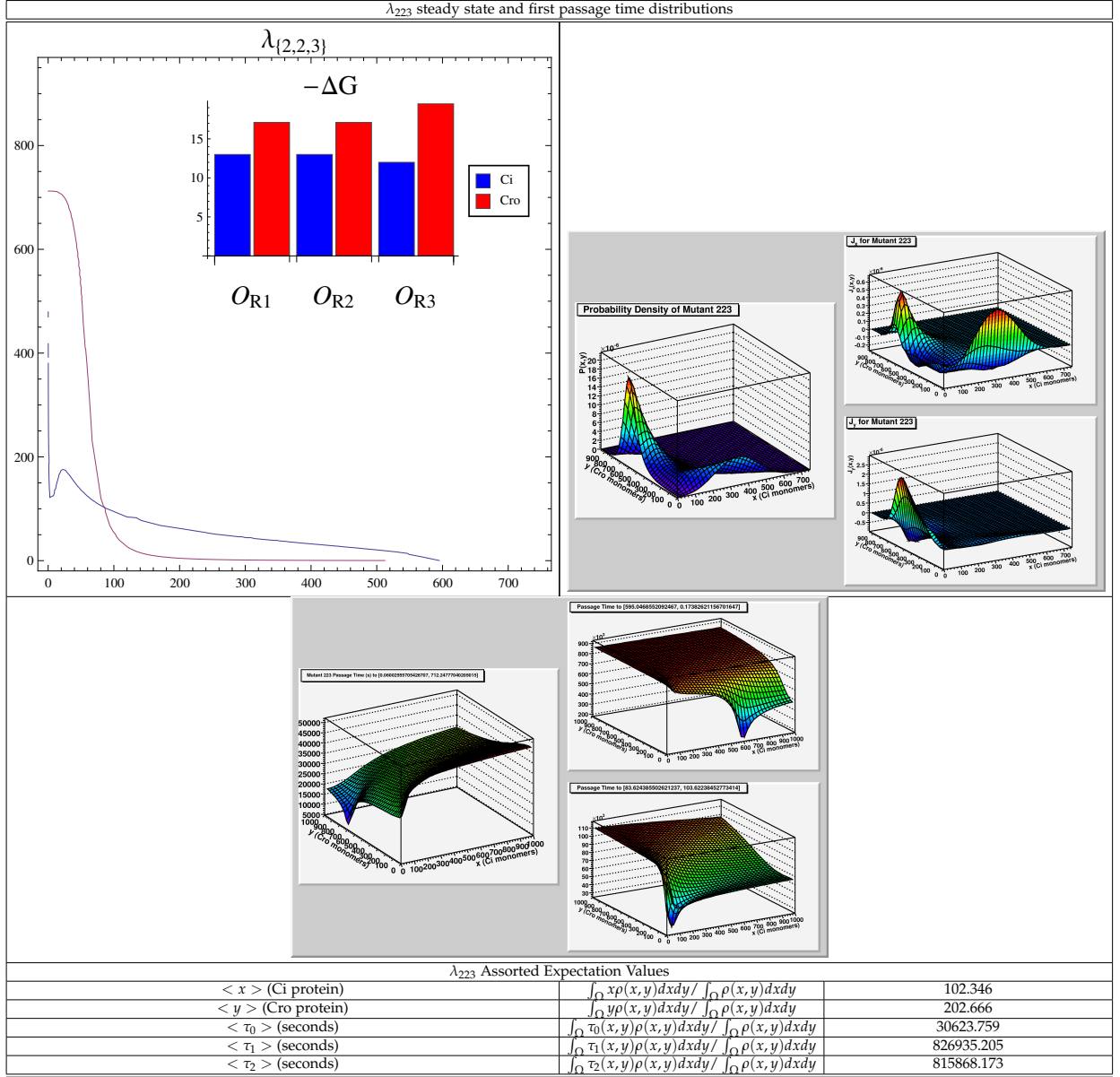
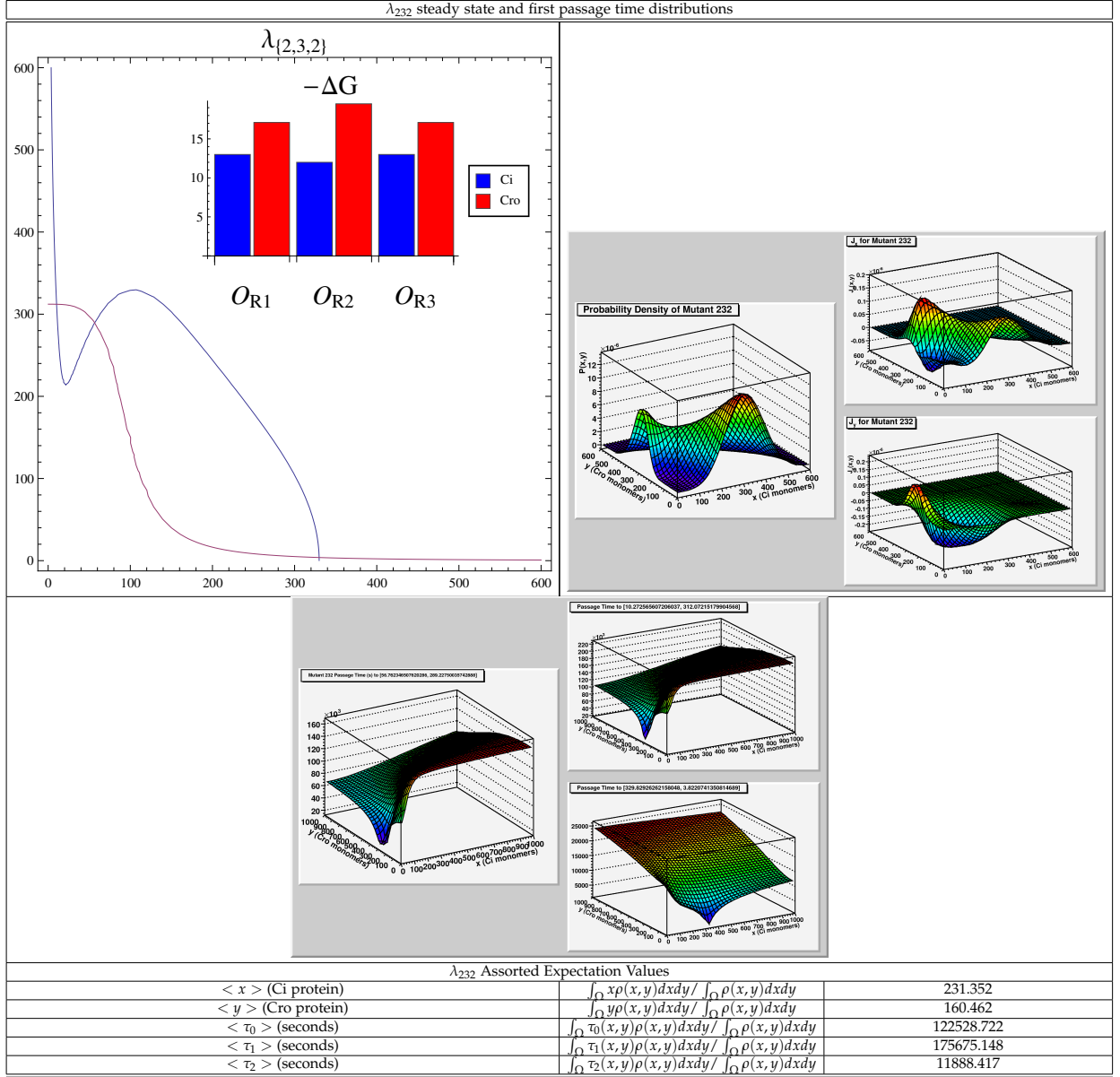
λ_{211} steady state and first passage time distributions			
O_{R1}	TAACACCGTGCGTGTG		
O_{R2}	TATCACCGCCAGAGGTA		
O_{R3}	TATCACCGCCAGAGGTA		
<div><div><div>$\lambda_{\{2,1,1\}}$</div><div></div></div><div><div><div>Probability Density of Mutant 211</div></div><div><div>J_1 for Mutant 211</div></div><div><div>J_2 for Mutant 211</div></div></div></div>			
<div><div><div>Mutant 211 Passage Time (s) to 177.80520012603306, 306.8769882329991</div></div><div><div>Passage Time to 327.30151486715674, 4.8304014577426792</div></div><div><div>Passage Time to 0.82445307384184452, 462.5709601591962</div></div></div>			
λ_{211} Assorted Expectation Values			
$\langle x \rangle$ (Ci protein)	$\frac{\int_{\Omega} x p(x,y) dx dy}{\int_{\Omega} p(x,y) dx dy}$	329.760	ratio to λ_{123} 3.543
$\langle y \rangle$ (Cro protein)	$\frac{\int_{\Omega} y p(x,y) dx dy}{\int_{\Omega} p(x,y) dx dy}$	120.570	0.710
$\langle \tau_0 \rangle$ (seconds to lytic)	$\frac{\int_{\Omega} \tau_0(x,y) p(x,y) dx dy}{\int_{\Omega} p(x,y) dx dy}$	7388.126	0.079
$\langle \tau_1 \rangle$ (seconds to ts)	$\frac{\int_{\Omega} \tau_1(x,y) p(x,y) dx dy}{\int_{\Omega} p(x,y) dx dy}$	39165.111	1.846
$\langle \tau_2 \rangle$ (seconds to lysogenic)	$\frac{\int_{\Omega} \tau_2(x,y) p(x,y) dx dy}{\int_{\Omega} p(x,y) dx dy}$	122448.847	0.464

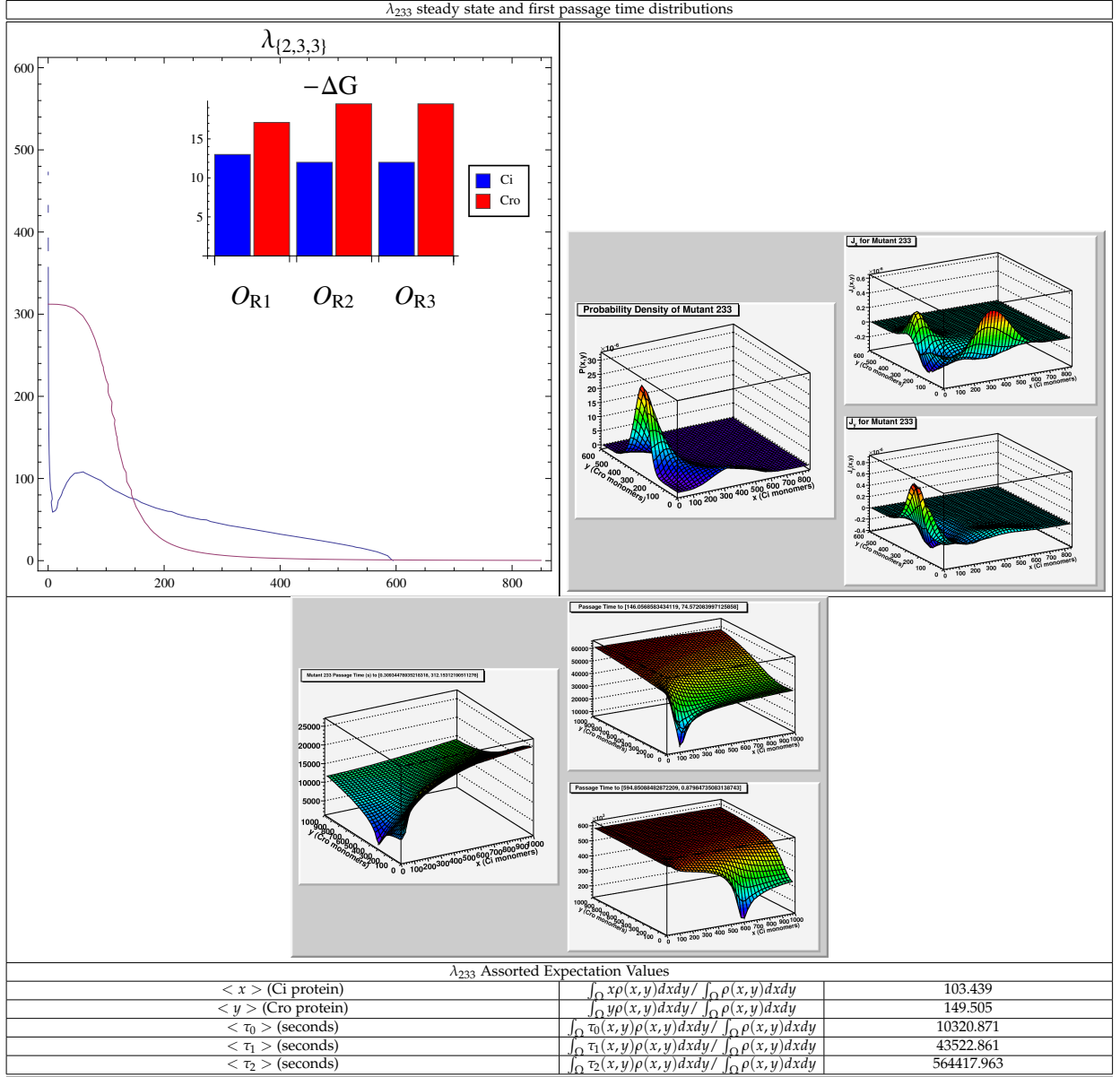
Table A.13: Properties of mutant λ_{211}

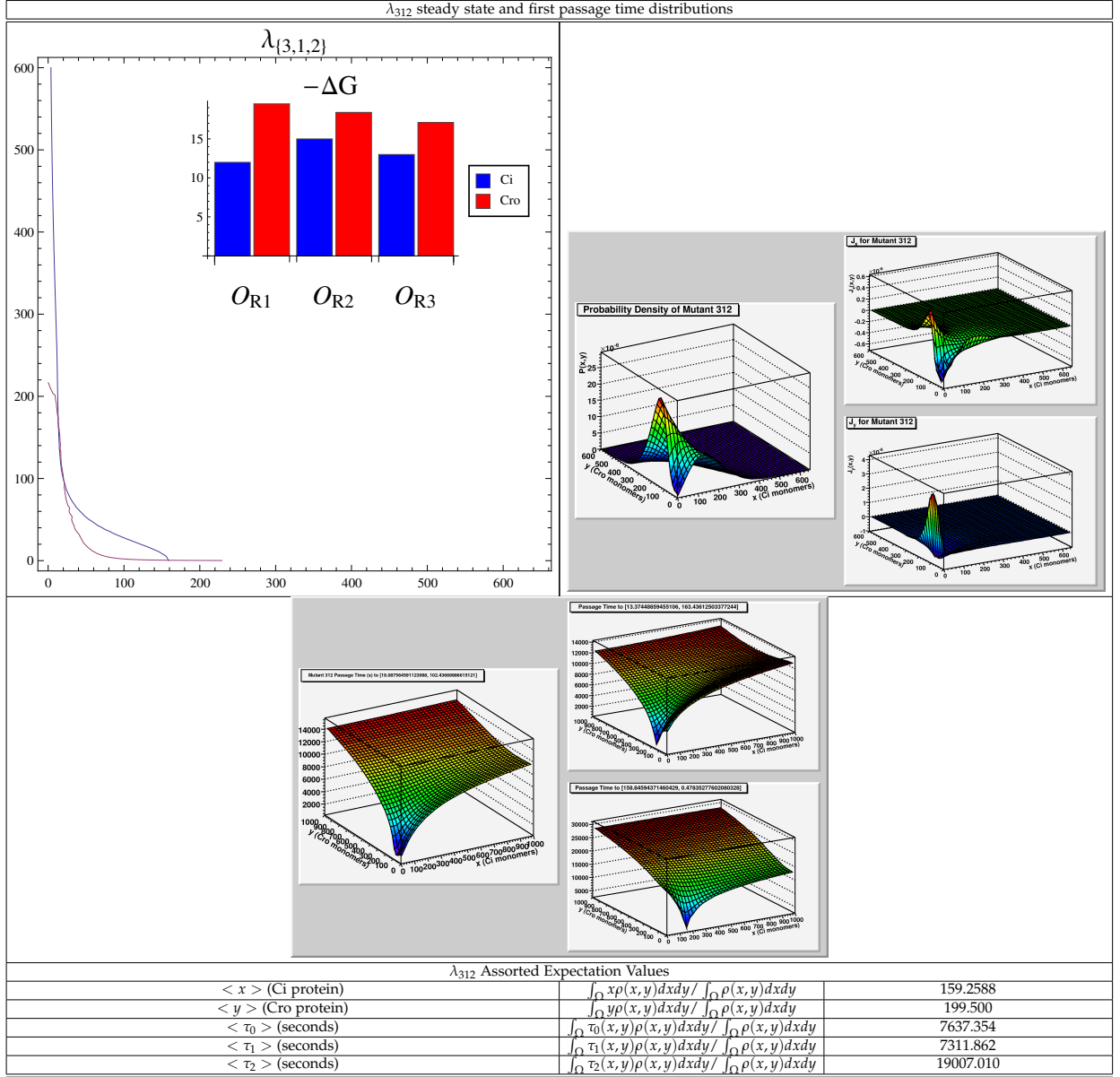
Table A.14: Properties of mutant λ_{213}

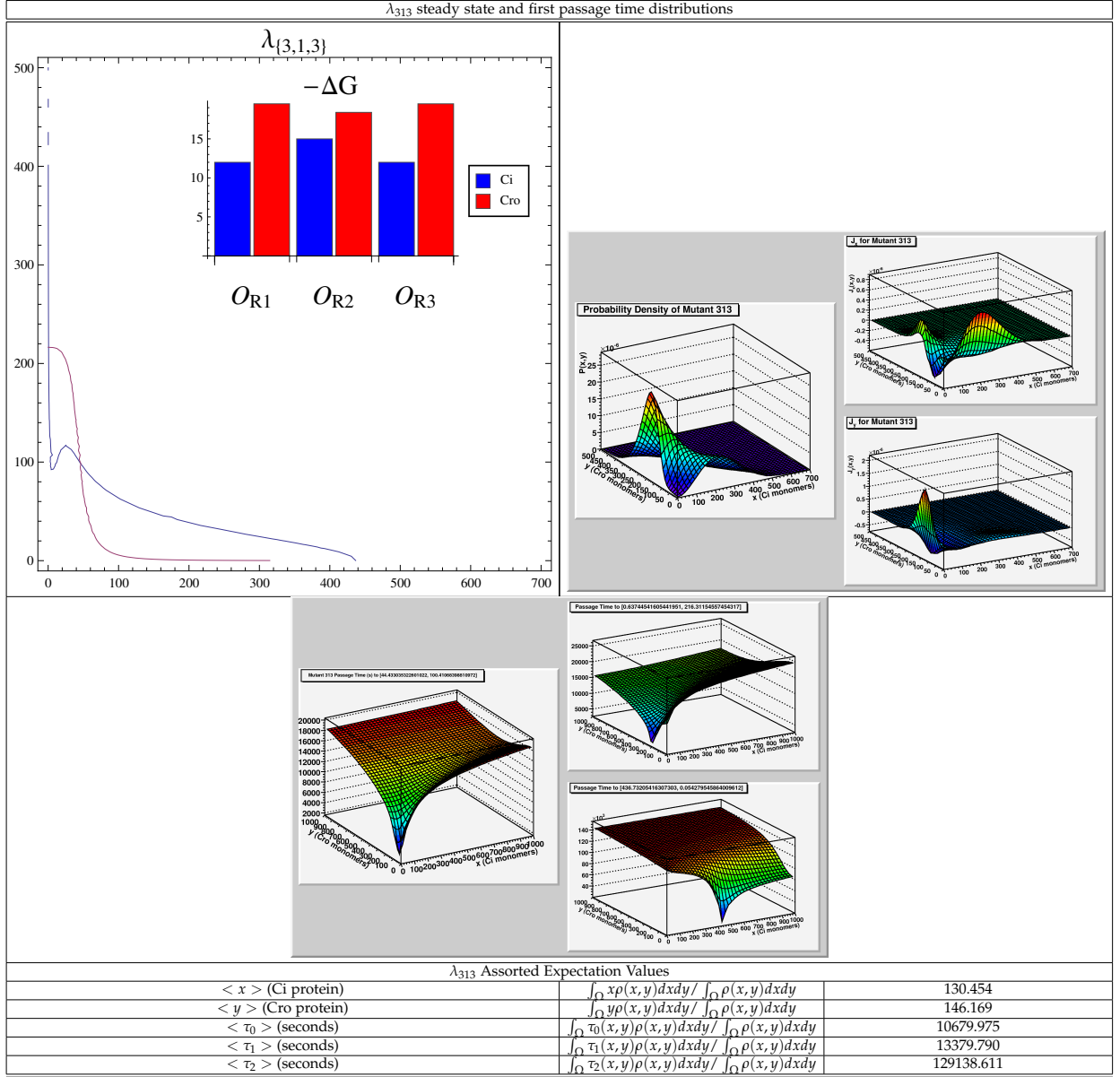
Table A.15: Properties of mutant λ_{222}

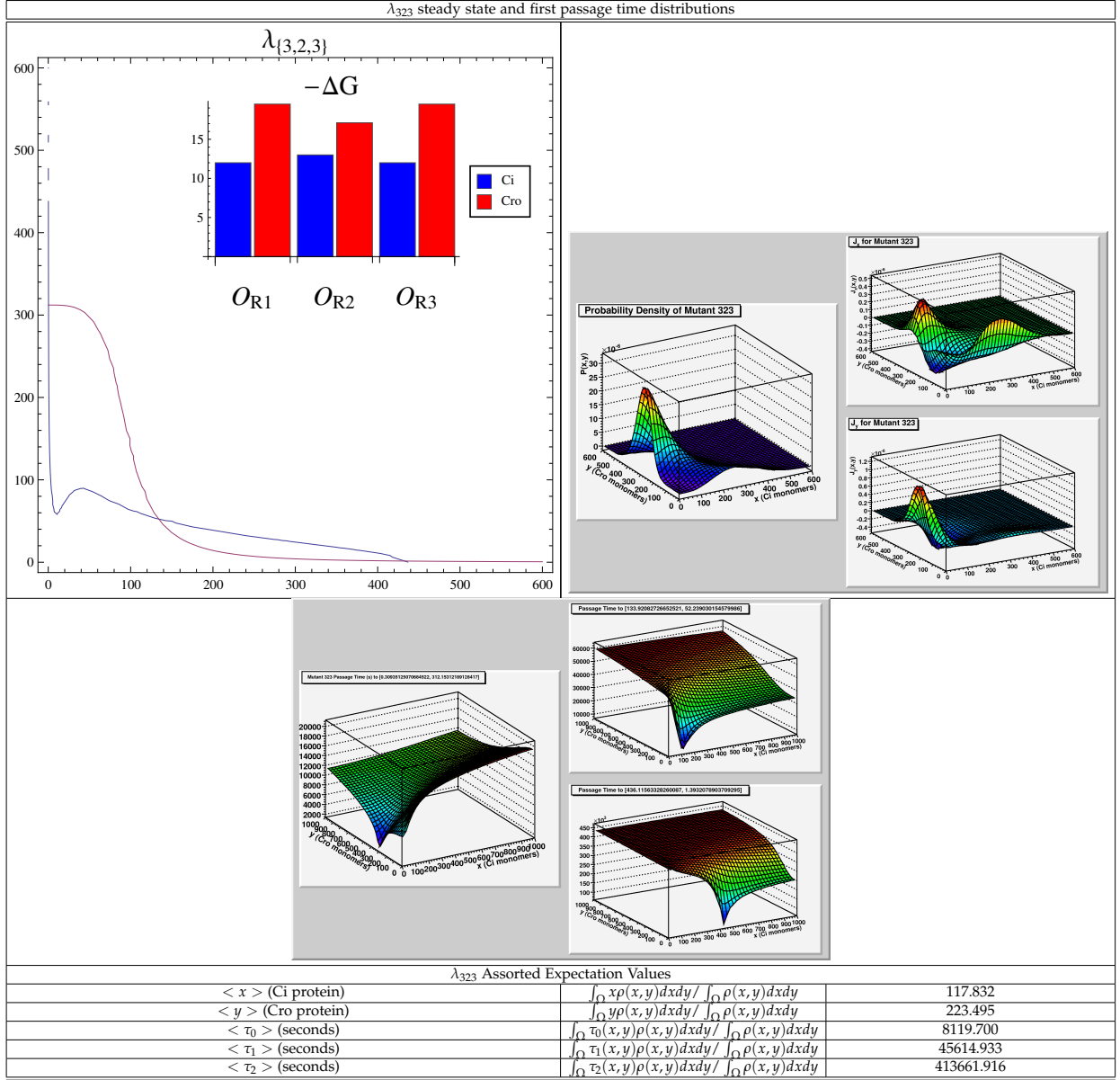
Table A.16: Properties of mutant λ_{223}

Table A.17: Properties of mutant λ_{232}

Table A.18: Properties of mutant λ_{233}

Table A.19: Properties of mutant λ_{312}

Table A.20: Properties of mutant λ_{313}

Table A.21: Properties of mutant λ_{323}

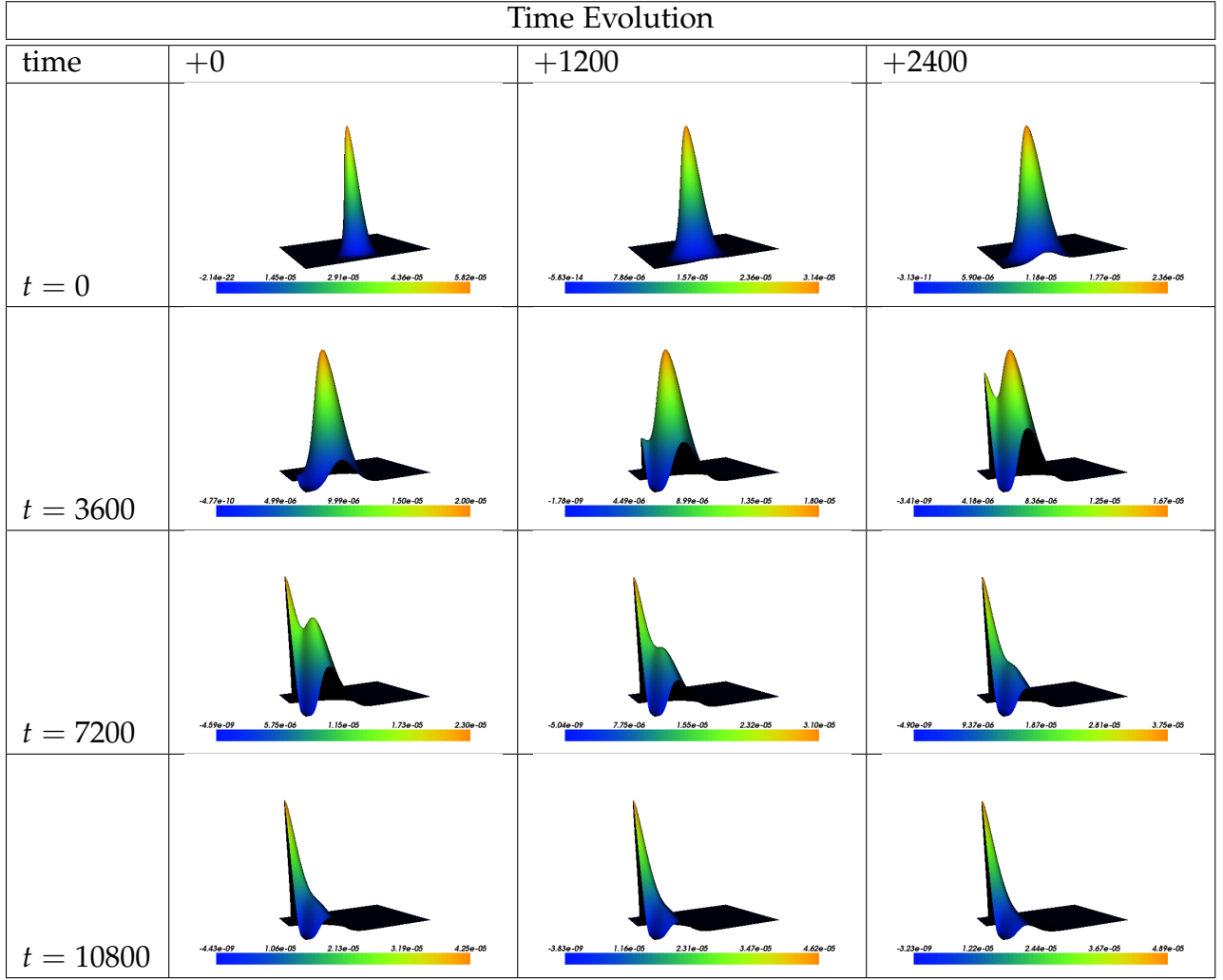


Table A.22: The time evolution of a gaussian initial condition relaxing into the steady state for the wild type with RNA polymerase in the partition function and constant diffusion tensor. Notice that the two states, lytic and lysogenic, are gapped only for the window of time from 4800 to 8400 seconds. Although the emphasis in the literature is with steady states, it is an interesting possibility that the two state phenomena is a strictly out of equilibrium phenomena, existing only for a window of time in the cell. A bacteria cell cycle is around 1800 seconds, suggesting that for many distributions the steady state is not reachable in the necessary time scale.

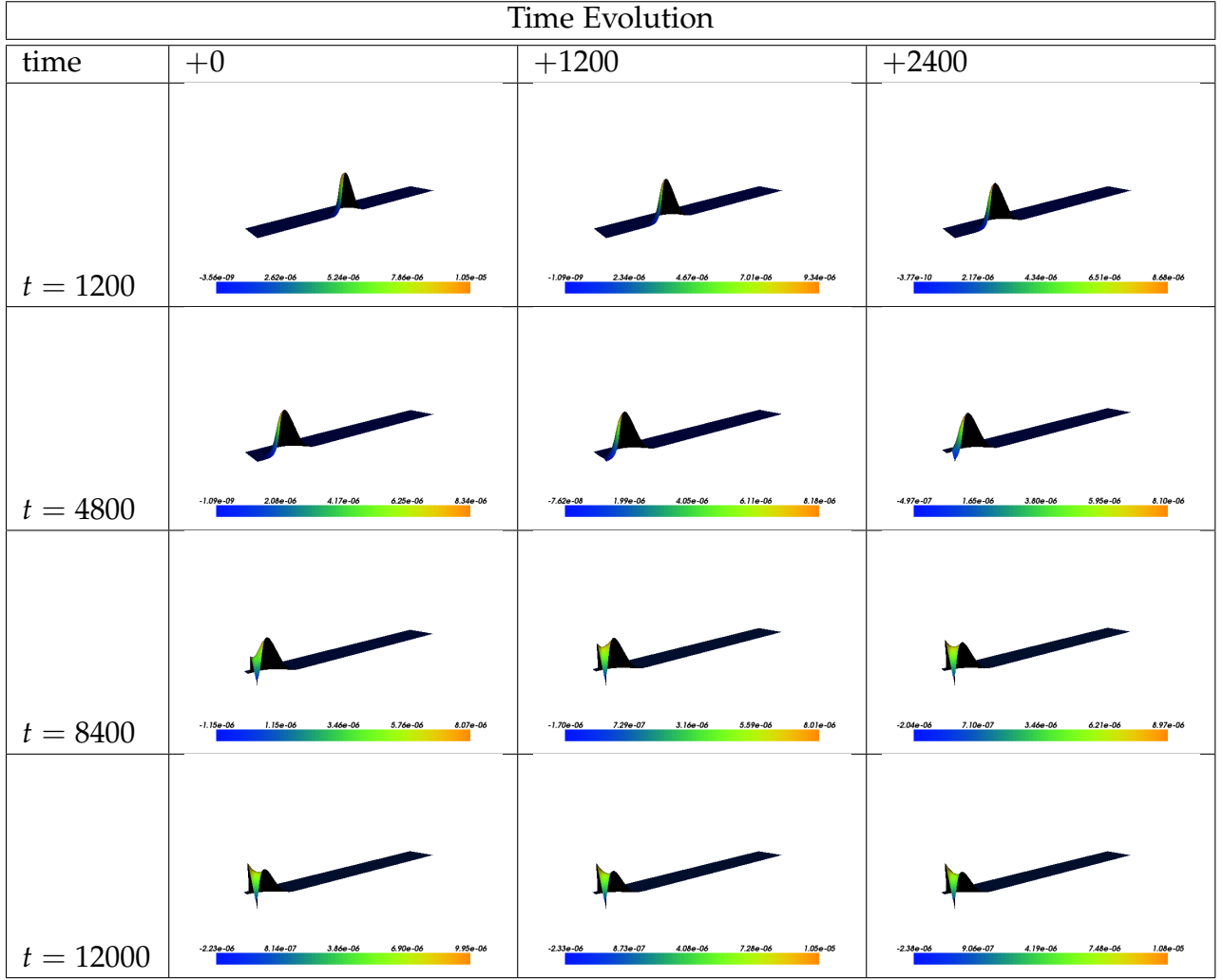


Table A.23: The time evolution of a gaussian initial condition relaxing into the steady state for the wild type with RNA polymerase in the partition function and constant diffusion tensor with different parameters from A.22 . The lysogenic peak in the classical dynamics is near 1500 cI proteins. To find a distribution with a second peak required this enlarged domain and was facilitated by an initial condition with many cI (3500). The short axis is 0-1000 Cro proteins, and the long axis is 0-5000 cI proteins. One can see the global effects through the large displacement of the lysogenic peak from its stationary point.

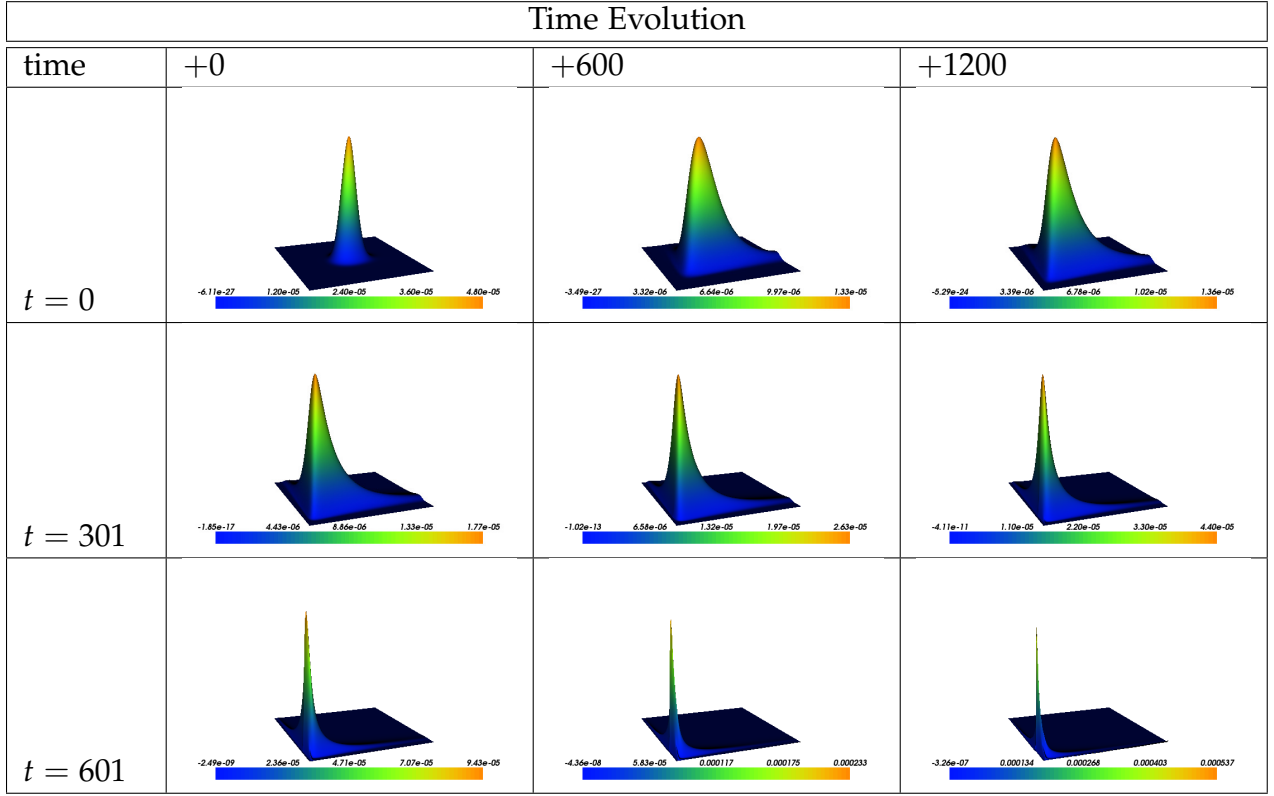


Table A.24: The time evolution of a gaussian initial condition relaxing into the steady state for the wild type with concentration-space dependant diffusion. The peak corresponds to the lytic state. Two strong tails develop, one extending towards potentially another peak.

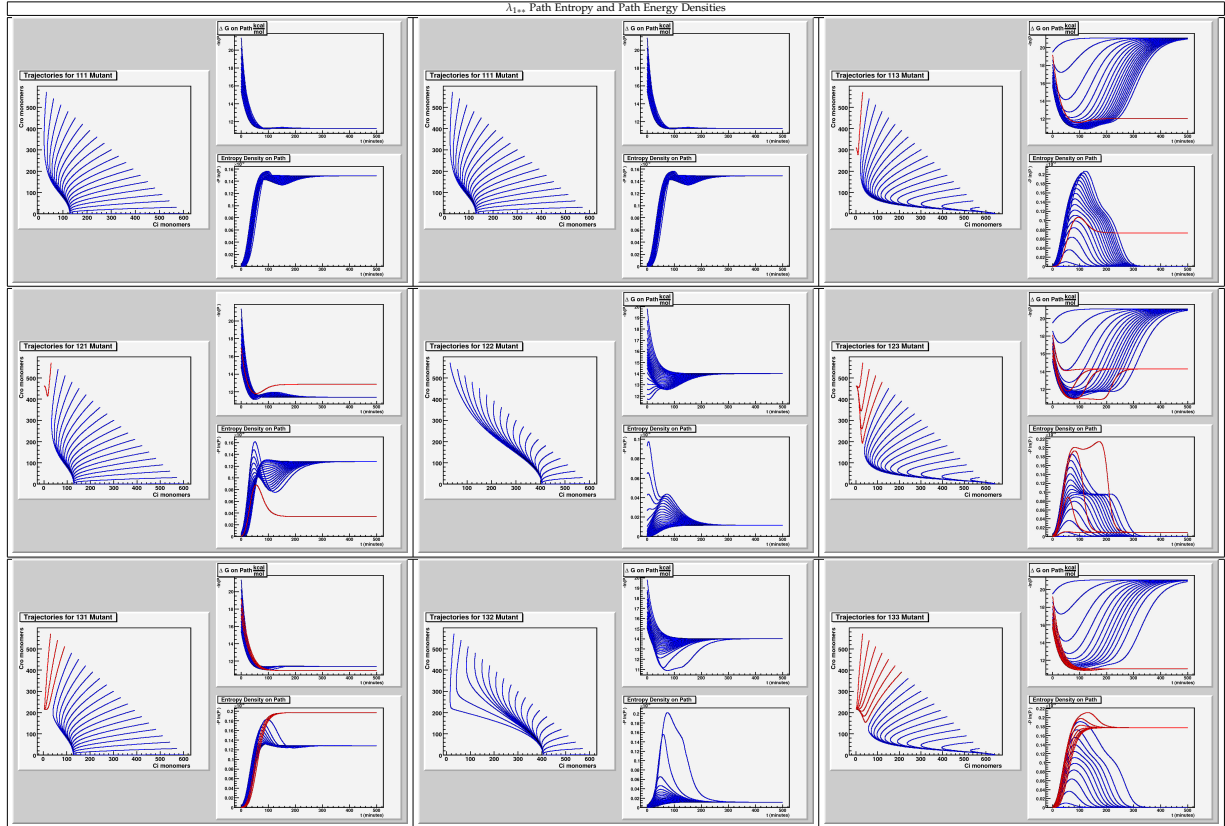


Table A.25: Entropy and Energy densities evaluated along example paths for mutations in $O_{R2,R3}$ with the O_{R1} fixed. Different trajectories can have different limiting values of entropy and energy density depending on the path taken.

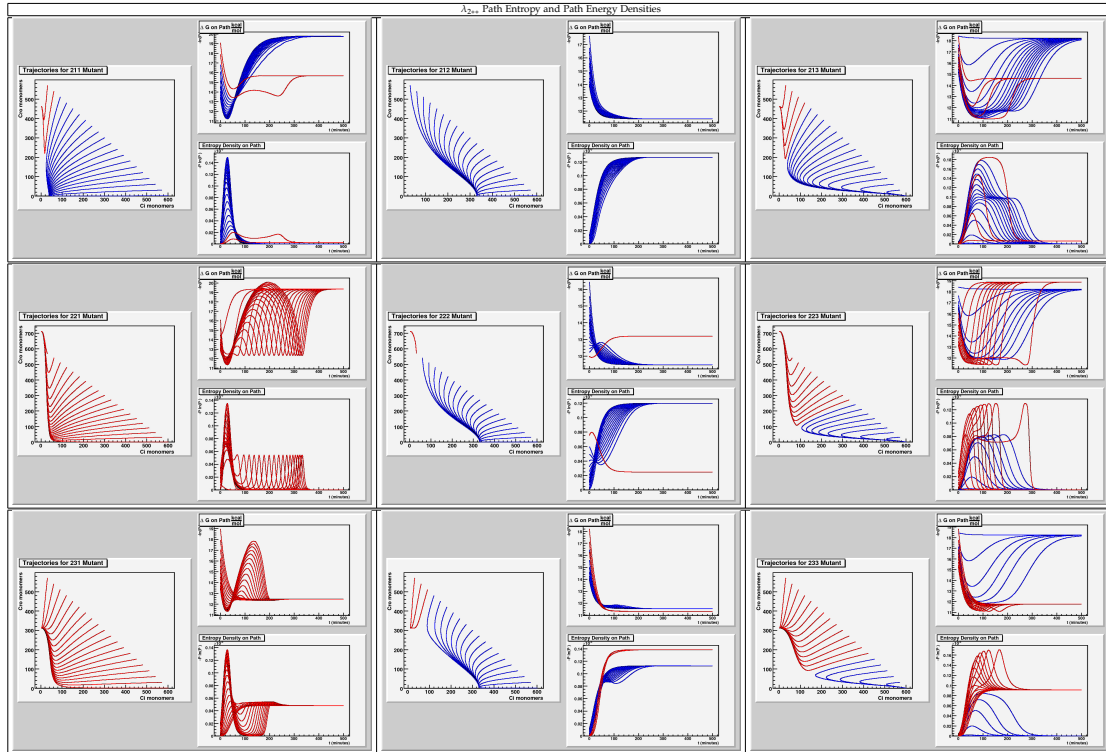


Table A.26: Entropy and Energy densities evaluated along example paths for mutations in $O_{R2,R3}$ with the O_{R1} mutated into O_{R2} . Different trajectories can have different limiting values of entropy and energy density depending on the path taken.

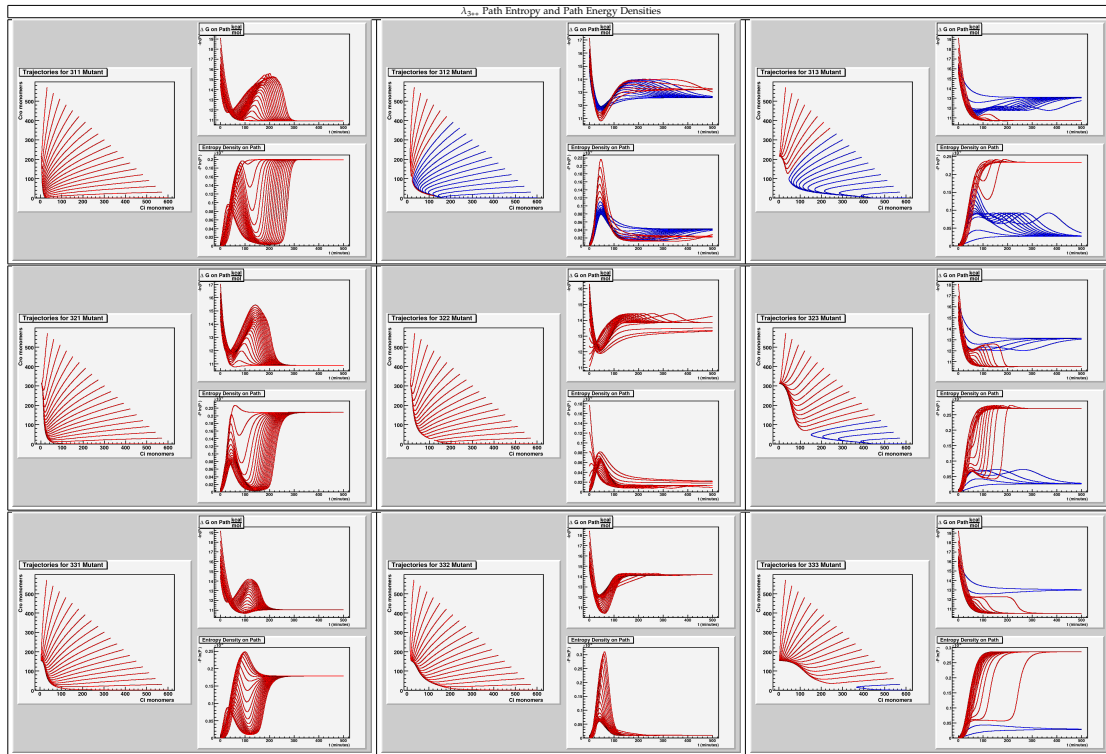


Table A.27: Entropy and Energy densities evaluated along example paths for mutations in $O_{R2,R3}$ with the O_{R1} mutated into O_{R3} . Different trajectories can have different limiting values of entropy and energy density depending on the path taken.

Sometimes the glass unbreaks and falls to heaven.
Quickly approaching the zero, *The Predicate*